THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES 1

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies
de l'Information et de la Communication
Spécialité : *Informatique*

Par

Yiru ZHANG

Modeling and management of imperfect preferences with the theory of belief functions

Thèse présentée et soutenue à Rennes, le 03/02/2020
Unité de recherche : IRISA
Thèse N° :

Rapporteurs avant soutenance :

Christine LARGERON	Professeure, Université Jean Monnet
Sébastien DESTERCKE	Chargé de recherche HDR CNRS, Université Technologie de Compiègne (UTC)

Composition du Jury :
Attention, en cas d'absence d'un des membres du Jury le jour de la soutenance, la composition du Jury ne comprend que les membres présents

Examinateurs :	Philippe BESNARD	Directeur de recherche CNRS, IRIT
	Meltem ÖZTÜRK	Maître de conférences, Université Paris Dauphine
	Sébastien DESTERCKE	Chargé de recherche HDR CNRS, UTC
	Christine LARGERON	Professeure, Université Jean Monnet
Dir. de thèse :	Arnaud MARTIN	Professeur, Université Rennes 1
Co-encadrante :	Tassadit BOUADI	Maître de conférences, Université Rennes 1

Invité(s) :

Acknowledgment

Being indebted to many people and organizations for their help and support during my Ph.D. study and research in team DRUID, IRISA, University of Rennes 1, I would like to express all my gratitude.

Before all, I would give my gratitude to the supervisors of this thesis, Arnaud Martin, professor of Université Rennes 1 and Tassadit Bouadi, Associated professor of Université Rennes 1, for their guide and help during my three years of thesis work as well as in the redaction of this thesis. Their patience, kindness and generosity have solved many problems both technically and emotionally. Their instructions also helped me to develop a rigorous way of research, which is a great treasure in my work and life in the future.

I would specially thank Yewan Wang, who have accompanied with me, during the moments of both depression and happiness. Her heterogeneous care and aids covered from my life to research work. The experiences of supporting each other on the redaction our own thesis at the same moment would be an unforgettable memory full of sweets and bitters. I would also promise her a loyal and promising future for both of us.

A great gratitude is also given to Mr. and Mrs. Pouliquen. They did not only rent the charm studio to me, but also treated me as their own grandson. I really wish them to keep in good health forever.

I will not forget the help from Mr. Yvan Picaud and Dr. Philippe Legay, my ex-colleagues in Orange Labs, Lannion. Without their introduction of this thesis subject, all this journey would not take place.

This dissertation would not have been possible without the support, contribution, and friendship of all my colleagues in team DRUID and my friends in Lannion. Na Li, Constance Thierry, Salma Ben Dhaou, ZhenZhe Zhong, Mouloud Kharoune and so many others. their presence, accompany, support during the days in Lannion have well decorated these three years of my life.

I am grateful for the ceaseless support and love of my parents and my brother. Their unwavering faith and confidence in me and my abilities have shaped me into the person I am. I will do my best to recompense for their supports and comprehensions on years of my absence from their side even though they will never be repayable.

I would firstly thank all members in my jury committee, especially Christine Largeron and Sébastien Desdercke, who have also been in the comité de suivi individuel of my thesis.

Finally, I would like to thank two entities, Allocations de Recherche Doctorale (ARED) and Lannion-Trégor Communauté, for their economical support on the work of my thesis.

Abstract

Preferences play an important role in daily life in various domains, especially in psychology, economy sociology and in philosophy. Today, the study of preferences has pivoted to unconventional preference models and new applications for preference management such as preference learning. Imperfect preference data needs to be reasoned by more complicated model and preference learning has found its indispensable position in applications of search engine sorting, recommendation systems, and social network analysis.

In this thesis, we review state-of-the-art methods on preference modeling, aggregation and preference learning. Based on the theory of belief functions (TBF), we propose a model for imperfect preferences with uncertainty, imprecision, called BFpref model. In TBF, a piece of knowledge with uncertainty and imprecision is called evidential. Thus, preferences in BFpref model are also referred to as evidential preferences.

BFpref model is also capable of expressing incompleteness through total ignorance in the framework of the TBF. With this model, relevant strategies are proposed to fuse multiple evidential preferences. In addition, a distance on imperfect preferences is introduced to take into account the four types of preference relationships differently. This distance is called Weighted Singleton Distance (WSD).

The unsupervised classification on evidential preferences with BFpref model is also studied by distinguishing between complete and incomplete preferences. Indeed, all existing work in learning on evidential objects are not theoretically convincing. In this thesis, an impossibility theorem on clustering over evidential bodies in the theory of belief functions is proposed and proved.

The following part gives more detailed contents of context and main contributions in this thesis.

Context

The thesis commences with an introduction of aggregation and learning tasks over imperfect preferences.

Aggregation of preferences

Most conventional preference studies focus on problems of preference aggregation, also known as social choice theory. In this process, several preferences are aggregated into one, usually considered for collective decision-making. In this thesis, we focus only on the problem of single-criteria decision making, where each item being compared is represented by a single criterion.

Preference learning

Nevertheless, the study of preferences is not limited to social choice theory. With the fashionable development of artificial intelligence (AI), preference reasoning is regarded as a particularly promising research direction for the AI community. Preference learning is primarily about inducing preference models from empirical data. The process of inducing preferences is also referred to as preference elicitation.

Preference learning techniques are widely used, from search engine services to the construction of recommendation systems. Searching for groups among agents according to their preferences is a fundamental step in preference determination, also called preference clustering.

New challenges: Imperfect preference data

Initiated by the development of digital technologies, in particular the Internet boom, more and more possibilities to collect and exploit preference data are becoming possible. Such a development has led to a multitude of new challenges and issues in preference research, both at the theoretical and application levels. Problems include, but are not limited to, the management of preference data imperfection and the corresponding applications.

In the context of this thesis, there are three main aspects to the imperfection of preference data: Preferences with uncertainty Preferences with imprecision Preferences with incompleteness (data missing)

Uncertainty in preferences refers to epistemic situations on the knowledge of preference information where preference relationships cannot be described by sound and reliable opinions. Imprecision in preferences refers to cases in which multiple preference relationships are possible, usually caused by a lack of knowledge or implicitness about preference information. The imprecision is generally caused by flaws in preference acquisition, or data sources, such as uncertain agent opinion, conflicts between multiple sources, and implicit information.

The theory of belief functions (TBF, also called Evidential theory or Dampster-Shaffer theory) is a formal framework for representing and reasoning information with uncertainty and imprecision by extending both the set-membership approach and probability theory. The term “evidential” is used in TBF to refer to imperfect information containing both “imprecision” and “uncertainty”.

Incompleteness of preferences refers to cases where the preference relations are not observed on all elements of a data space. i.e. preference information is partially and/or completely missing.

Objectives

The objectives of this thesis are to propose a new modeling of imperfect preferences and to explore the problems of managing such preferences. To deal with the problems of uncertainty and imprecision, we rely on a theoretical framework: the theory of belief functions.

The subject of this thesis is globally related to three aspects: the modeling of imperfect preferences (with the theory of belief functions), the decision making on imperfect preferences (resulting from the aggregation of preferences), and the unsupervised learning of imperfect preferences.

Chapter 2 to Chapter 4 provide preliminary knowledge and review the state-of-the-art works on the corresponding topics.

In Chapter 2, different orders, structures, and relationships of preferences are presented and compared, followed by a synthesis on imperfect preference models, in particular the fuzzy preference model, which is the most popular one. Concerning the theory of belief functions, basic definitions as well as combination rules and their conditions of use are introduced.

In Chapter 3, existing similarity measures on preferences are introduced. These measures are divided into two categories: measures of pairwise preference relations and measures of preference structures. The first category comprises the standardized distances on preference encoding and the axiomatic distances. We also compared the different axioms accepted for each axiomatic distance and analyzed their differences in the interpretation of incomparability. The second category concerns measurements between two preference structures (or orders of preference). In this case, similarity is often measured by correlation-related distances such as Kendall's, Spearman's or Pearson's distance. As for the theory of belief functions, measures of similarity between BBAs are introduced in three categories: distances from a geometric point of view, divergence from a statistical distribution point of view, and conflict from the view of a common support on events.

Chapter 4 reviews major problems and techniques in preference aggregation and preference clustering. Particularly, techniques managing incomplete data are also reviewed. Concerning the aggregation of preferences, different classical voting methods are presented, followed by voting issues such as Condorcet's paradox. The relationship between similarity measures and preference aggregation is also reviewed. Indeed, for the majority of aggregation rules, in particular consensus rules, the aggregated result is essentially a preference closest to all the preferences to be aggregated, i.e. it is a process of minimizing the sum of the distances of the aggregated (group) preference to the preferences of the agents considered. The differences between the different methods of voting or aggregating preferences essentially correspond to differences in the distance applied.

Preference learning is a sub-domain of artificial intelligence, mainly consisting of two objectives (i) to classify agents into different groups based on their individual preferences, and (ii) to explicit group preferences. The first objective is often seen as a problem close to the detection of communities, and the second as a problem close to recommendation systems. In this thesis, the first objective is principally targeted.

Incompleteness of data bring challenges to both preference aggregation and learning. Techniques dealing with incompleteness consists of three principal categories:

- Discarding the missing data,
- Imputing the missing data,
- Modeling missing data by soft computing methods.

The adaptable cases as well as advantages/disadvantages are also concluded in Chapter 4.

Contributions

The contributions of this thesis are composed of three aspects: modeling, aggregation and learning of imperfect preference data, respectively introduced in Chapter 5, 6 and 7.

In preference modeling

Contributions in preference modeling are introduced in chapter 5. A novel model for uncertainty and imprecise preference is proposed, based on the theory of belief function, namely BFpref model (Belief Function based preference). An ambiguity in the definitions of preference relationships over the interpretations of "incomparabilit" and "incompleteness" is firstly pointed out. Most work interprets the "incomparability" relationship as missing or "undecided" information. Other work interprets "incomparability" as another type of preference relationship, different from "strict preferences" and "indifference", or the "undecidable" case, which also meets the original definition of "incomparability".

In this work, we consider that the incompleteness of preferences is caused by missing information and we have clarified this ambiguity by proposing BFpref model. According to the theory of belief functions, evidential bodies with uncertainty and imprecision are represented by BBAs under a predefined discrimination framework. By defining a framework of discernment about possible events (called singleton in TBF):

$$\Omega = \{\omega_1, \omega_2, ..., \omega_k\}$$

, the degrees of uncertainty and imprecision are represented by the Basic Belief Assignment (BBA) $m : 2^{\Omega} \leftarrow [0.1]$ such that :

$$\sum_{X \subseteq \Omega} m(X) = 1$$

Non-zero elements are called focal elements.

Uncertainty is represented by values less than 1 and imprecision by values on elements representing unions of events. For example, $1 > m(X_1 \cup X_2) > 0$ represents an uncertain degree of belief about imprecise information between X_1 and X_2.

The BFpref model allows the user to express preferences based on a degree of belief. These degrees are defined on each pair alternatives, and are interpreted as elementary belief masses. The discernment frame of BFpref model is composed of four exclusive singletons, representing respectively the relations "strict preference", "strict inverse preference", "indifference" and "incomparability". Formally, in BFpref, the discrimination framework is defined as follows: $\Omega = \{\omega^{\succ}, \omega^{\prec}, \omega^{\sim}\}$.

The BFpref model is capable of expressing all three aspects of imperfection in preferences, and it distinguishes ambiguity from incomparability. The "undecidable" case is directly represented by a singleton, while missing information ("undecided" case) is represented by a case of extreme imprecision - total ignorance, which is the union of all possible singletons on the framework of discernment.

In preference aggregation

Contributions in preference aggregation is introduced in Chapter 6. A preference aggregation strategy is proposed along with a novel distance over BBAs for decision making. With the BFpref model, a preference aggregation strategy is proposed, based on the Dempster combination rule and a minimum distance decision strategy. We also proposed a strategy to avoid Condorcet's paradox and an efficient Depth First Search (DFS) method for constructing acyclic oriented graphs. Compared to a naive algorithm, the proposed DFS method improves results on most graph structures.

During the decision stage of the preference aggregation process modeled by BFpref, a flaw in the distances for evidential bodies is identified: all existing distances for BBAs consider that the distances between singletons are identical, differences in the similarity between singletons are never dealt with. Indeed, by measuring the similarity between preference relationships, this problem is particularly important. For example, the similarity between "difference" and "strict preference" is equivalent to the similarity between "strict preference" and "reverse strict preference". This result obtained with all existing distances is counter-intuitive, as the first distance is expected to be less than the second. As BBAs are in space of 2^{Ω}, the interaction between focal elements should also be considered. This property is called strongly structural.

To solve this problem, we analyzed the assumptions applied in the Jousselme distance, which is a strongly structural distance for BBAs, and removed an unwanted assumption that the distances between the different singletons are equal. By extending the Jousselme distance, we proposed the Weighted singleton distance (WSD). The

WSD distance takes into account the differences in similarities between the different singletons, and is furthermore highly structural. In other words, the measure between singletons' unions also considers different similarities between singletons. We also applied this WSD distance in the decision making stage of the SUSHI dataset (A Survey of Sushi Preferences in Japan). The comparison results show that decisions based on the WSD are more reasonable and moderate than those based on the Jousselme distance.

In preference learning

Contributions in preference learning is introduced in Chapter 7, mainly consisting of a strategy for imperfect preference reasoning and clustering, as well as an impossibility theorem for clustering over evidential bodies.

Learning on evidential preferences, more precisely, the clustering of preferences from the BFpref model, is also addressed in this thesis. Agents' profiles are represented by uncertain preferences via the BFpref model when several sources of conflicting preferences are taken into account. In our method, the estimation of BBAs of an identical agent corresponds to the estimation of the conflict between the different sources of information. More precisely, the dissimilarities between the different sources of an identical agent are considered as a degree of ignorance of this agent, and the relations represented for each source are distributed over the corresponding elements of the BBAs. The similarities between the agents are based on the sum of the Jousselme distance over all pairs of objects studied, which is equivalent to the Kendall distance in the case of definite preferences.

Concerning the classification phase, the k-centroid method is not applicable because only the distances between couples are provided. Therefore, the independent algorithms of the centroid computation are the only ones that can be applied in our case. In this work, the Ek-NNclus (Evidential k-NN clustering) method is applied. Ek-NNclus is a flexible clustering method also based on belief function theory. The result of the clustering carried are evaluated by the silhouette score and Adjusted Rand Index (ARI) in comparison with other metrics such as Kendall distance and Euclidean distance.

By comparing with a strategy applying an arithmetic mean of the two sources of conflicting agent preferences, it is illustrated that the BFpref model returns a better clustering result. Our method is therefore able to detect agent communities even in the presence of uncertain preferences.

Clustering has so far only been performed on complete preferences. Indeed, the distance for incomplete evidential orders based on the existing distances does not allow the classification of incomplete data. This is caused by ignorance (partial or total) expressed in the BBAs. In the frame of TBF, "ignorance" is considered as an extreme case of imprecision. The imprecision is fully considered in conjunctive combination rules in TBF, but not in distance measuring over BBAs. An obvious consequence is that all "ignorance" are considered as identical. Thus, evidential objects with high imprecision are usually grouped into one cluster. Such flaw in the distance seriously jeopardizes the learning task.

We launch a discussion on learning tasks over evidential bodies, started by the relation between k-centroid and combination rules in TBF. An impossibility theorem for k-centroid clustering is proposed and proved, stating that it is impossible to simultaneously satisfy the properties of “metric consistency, “surjectivity of centroid” and “neutrality of ignorance”. A corresponding corollary is also induced, stating that no conjunctive combination rule is suitable for k-centroid clustering methods over BBAs.

Perspectives

Other than the contributions introduced above, in thesis, we have also proposed a preliminary idea of clustering evidential bodies by abandoning the property of "metric consistency". We reckon that the similarity between evidential bodies should also be measured by evidential values in another discernment frame of TBF. This idea is expected to be explored and verified, which stays in our short-term perspectives.
Furthermore, the decision making problem is limited in mono-criteria in this thesis. Multi-criteria decision making (MCDM) is a more commonly encountered problem. PROMETHEE is a famous MCDM method, and is also in frame of pairwise comparison among alternatives as in BFpref model. Thus, extending BFpref on PROMETHEE for MCDM problems would be a promising perspective.

Résumé

Les préférences jouent un rôle important dans notre vie quotidienne, et ce dans diverses domaines, notamment en psychologie, en économie, en sociologie et en philosophie. L'étude des préférences a une longue histoire. Cela a été mis en lumière par les chercheurs pendant des siècles, en particulier dans les systèmes de vote. De nos jours, l'étude et la modélisation des préférences s'est orientée vers dee modèles des préférences non conventionnels, et vers de nouvelles applications concernant la gestion de ces dernières. La théorie des probabilités, des sous-ensembles flous, des ensembles aléatoires, etc., ont été introduites pour modéliser les préférences incertaines. Aujourd'hui, de nombreux travaux de recherche sur les préférences portent sur des applications de tris dans les moteurs de recherche, de systèmes de recommandation, et d'analyse de réseaux sociaux.

Dans cette thèse, nous passons en revue les méthodes de pointe sur la modélisation des préférences, l'agrégation et l'apprentissage des préférences. Basé sur la théorie des fonctions de croyance (TBF), nous proposons un modèle pour les préférences imparfaites avec incertitude, imprécision, appelé modèle BFpref. Dans la TBF, un élément de connaissance avec incertitude et imprécision est appelé évidentiel. Ainsi, les préférences dans le modèle BFpref sont également appelées préférences évidentielles.

Dans le cadre du TBF, le modèle BFpref est également capable d'exprimer l'incomplétude par une ignorance totale. Avec ce modèle, des stratégies pertinentes sont proposées pour fusionner de multiples préférences probantes. En outre, une distance sur les préférences imparfaites est introduite pour prendre en compte différemment les quatre types de relations de préférence. Cette distance est appelée distance singleton pondérée (Weighted Singleton Distance, WSD).

La classification non supervisée sur les préférences évidentielles avec le modèle BFpref est également étudiée en distinguant les préférences complètes et incomplètes. En effet, tous les travaux existants en matière d'apprentissage sur les objets probants ne sont pas théoriquement convaincants. Dans cette thèse, un théorème d'impossibilité de clustering sur les corps évidentiels dans la théorie des fonctions de croyance est proposé et prouvé.

La partie suivante donne un contenu plus détaillé du contexte et des principales contributions de cette thèse.

Contexte

La thèse commence par une introduction des tâches d'agrégation et d'apprentissage sur les préférences imparfaites.

Agrégation des préférences

La plupart des études classiques sur les préférences porte sur des problèmes d'agrégation des préférences, ce qui est également appelé théorie du choix social. Dans ce processus, plusieurs préférences sont agrégées en une seule, généralement considérée pour une prise de décision collective. Dans cette thèse, nous nous concentrons uniquement sur le problème de prise de décision monocritère, où chaque élément comparé est représenté par un seul critère.

Apprentissage des préférences

Néanmoins, l'étude des préférences n'est pas limitée à la théorie du choix social. Avec le développement en vogue de l'intelligence artificielle (IA), le raisonnement avec les préférences est considéré comme une direction de recherche particulièrement prometteuse pour la communauté de l'IA. L'apprentissage de préférences s'agit principalement d'induire des modèles de préférence prédictifs à partir de données empiriques. Le processus d'induction de préférences prédictives est également appelé élicitation de préférences.

Les techniques d'apprentissage des préférences sont très utilisées, des services de moteur de recherche à la construction de systèmes de recommandation. La recherche de groupes parmi les agents en fonction de leurs préférences est une étape fondamentale de la détermination des préférences, également appelée clustering de préférences.

Nouveaux défis: imperfection des données de préférence des agents

Initié par le développement des technologies numériques, en particulier le boom d'Internet, de plus en plus de possibilités pour collecter et exploiter les données de préférences deviennent possibles. Un tel développement a entraîné une multitude de défis et d'enjeux nouveaux en matière d'étude des préférences, tant au niveau théorique qu'au niveau applicatif. Les problèmes incluent mais ne se limitent pas à la gestion de l'imperfection des données de préférence et aux applications correspondantes.

Dans le cadre de cette thèse, l'imperfection des données issues des préférences comporte principalement trois aspects :

- Les préférences incertaines
- Les préférences imprécises
- Les préférences incomplètes

L'incertitude dans les préférences fait référence aux situations épistémiques sur la connaissance des informations de préférences où les relations de préférences ne peuvent pas être décrites par un avis solide et sûr. L'imprécision dans les préférences fait référence aux cas dans lesquels de multiples relations de préférence sont possibles, généralement causées par l'absence de connaissances ou l'implicite sur les informations relatives aux préférences. Les imperfections sont généralement causées par les failles dans l'acquisition des préférences, ou les sources des données, telles que l'opinion incertaine des agents, les conflits entre plusieurs sources, et les informations implicites.

La théorie des fonctions de croyance est un cadre formel permettant de représenter et de raisonner avec des informations incertaines et imprécises en élargissant à la fois l'approche fondée sur l'appartenance à un ensemble et à la théorie des probabilités. Le terme « crédibiliste » est utilisé dans la théorie de fonction de croyance pour se référer à des informations contenant à la fois « l'imprécision » et « l'incertitude ».

L'incomplétude des préférences renvoie aux cas où la relation de préférence n'est pas observée sur tous les éléments d'un espace de données, i.e. l'information sur les préférences est partiellement et/ou complètement manquante.

Objectifs

Les objectifs de cette thèse sont de proposer une nouvelle modélisation des préférences imparfaites et d'explorer les problèmes de gestion de ce type de préférences. Pour traiter les problèmes d'incertitude et d'imprécision, nous nous sommes appuyé sur un cadre théorique solide : la théorie des fonctions de croyance.

Le sujet de cette thèse concerne globalement trois aspects : la modélisation des préférences imparfaites (avec la théorie des fonctions de croyance), la prise de décision sur de telles préférences (issues de l'agrégation des préférences), et l'apprentissage non-supervisé sur de telles préférences (clustering des préférences imparfaites).

Cette thèse est principalement structurée par une partie état-de-l'art et une partie présentant les contributions. La partie sur l'état-de-l'art présente trois chapitres :

- Les concepts de base autour des préférences et la théorie des fonctions de croyance
- Les mesures de similarité sur les préférences et les fonctions de croyance
- La gestion des préférences, incluant l'agrégation des préférences et l'apprentissage des préférences

Dans Chapitre 1, différents ordres, structures, et relations de préférences sont présentés et comparés, suivi par une synthèse sur les modèles des préférences imparfaites, en particulier le modèle des préférences floues, qui est le plus populaire. Concernant la théorie des fonctions de croyance, les définitions de base ainsi que les règles de combinaison et leurs conditions d'utilisation sont introduits.

Dans le deuxième chapitre, des mesures de similarité sur des préférences sont d'abord introduites. Ces mesures sont divisées en deux catégories : les mesures sur les relations de préférences par paire et celles sur les structures des préférences. La première catégorie comporte les distances normées sur l'encodage des préférences et les distances axiomatiques. Nous avons aussi comparé les différents axiomes acceptés pour chaque distance axiomatique et analysé leurs différences d'interprétation de l'incomparabilité. La seconde catégorie concerne les mesures entre deux structures de préférence (ou ordres de préférence). Dans ce cas, la similarité est souvent mesurée par des distances liées à la corrélation telle que la distance de Kendall, de Spearman ou de Pearson. Quant à la théorie des fonctions de croyance, les mesures de similarité entre les fonctions de masse sont introduites en trois catégories : les distances d'un point de vue géométrique, la divergence d'un point de vue de distribution statistique et le conflit d'un point de vue de support sur les évènements.

Chapitre 3 concerne principalement l'agrégation et l'apprentissage des préférences, et surtout le clustering des préférences. Concernant l'agrégation des préférences, différentes méthodes classiques de votes sont présentées, suivie par des problématiques de vote telle que le paradoxe de Condorcet et le théorème d'impossibilité d'Arrow. La relation entre les mesures de similarité et l'agrégation de préférences est aussi passée en revue. En effet, pour la majorité des règles d'agrégation, en particulier les règles de consensus, le résultat d'agrégation est essentiellement une préférence la plus proche de toutes les préférences à agréger, i.e. c'est un processus de minimisation de la somme des distances de la préférence agrégée (de groupe) aux préférences des agents considérées. Les différences entre les différentes méthodes de vote ou d'agrégation de préférences correspondent essentiellement à des différences au niveau de la distance appliquée.

Nous avons également examiné les méthodes d'agrégation des préférences floues, dont la majorité est fondée sur les opérateurs de moyenne pondérée (Operator of weighted average, OWA en anglais). Par contre, ce modèle est incapable d'exprimer l'imprécision dans les préférences.

L'apprentissage des préférences est un sous-domaine de l'intelligence artificielle. Dans cette thèse, nous nous focalisons principalement sur la clustering des préférences en considérant essentiellement les problèmes liés à l'incomplétude des données, car il s'agit d'une problématique souvent rencontrée dans le contexte les différentes applications utilisant ou manipulant des préférences réelles (i.e. non simulées). Nous avons étudié les méthodes de clustering les plus utilisées pour les données manquantes et les avons catégorisé en trois groupes :

- Défausser des données manquantes
- Déduction des données manquantes
- Modéliser les données manquantes en modélisant les imperfections.

Quant à l'apprentissage des préférences, deux objectifs sont considérés dans le cadre de cette thèse :

- classifier les agents en différents groupes en se basant sur leurs préférences individuelles, et
- expliciter les préférences de groupes. Le premier objectif est souvent vu comme un problème proche de la détection des communautés, et le second comme une problématique proche des systèmes de recommandation.

Contributions

Les contributions de cette thèse sont composées de trois aspects : la modélisation, l'agrégation et l'apprentissage des données de préférences imparfaites, respectivement introduits dans les chapitres 5, 6 et 7.

En modelisation des préférences

Premièrement, nous avons souligné l'ambiguïté sur les définitions de la relation de préférence « incomparabilité » et « l'incomplétude ». La plupart des travaux interprètent « incomparabilité » comme une information manquante ou « non décidé ». D'autres travaux interprètent « incomparabilité » comme une relation binaire spécifique, différente des « préférences strictes » et « d'indifférence », ou du cas « non décisif », qui respecte également la définition originale de « l'incomparabilité ».

Dans ce travail, nous considérons que le caractère incomplet des préférences est causé par les informations manquantes et nous avons clarifié cette ambiguïté dans un nouveau modèle de préférences fondées sur la théorie des fonctions de croyance, nommé modèle BFpref. Selons la théorie des fonctions de croyance, l'incertitude et l'imprécision sont représentées par les fonctions de masse m() sous une cadre de discernement prédéfinie. En définissant un cadre de discernement sur les évènements possibles:

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_k\},$$

les degrés d'incertitude et d'imprécision sont représenté par la fonction de masse m : $2^\Omega \leftarrow [0, 1]$ tel que

$$\sum_{X \subseteq \Omega} m(X) = 1$$

où l'imprécision est représenté par les valeur sur les non-singleton éléments.

Le modèle BFpref est un modèle de préférences par paires dont le cadre de discernement est composé de quatre singletons, représentant respectivement les termes « préférence stricte », « préférence stricte inverse », « indifférence » et « incomparabilité ». Formellement, dans BFpref, le cadre de discernement:

$$\Omega = \omega^{\succ}, \omega^{\prec}, \omega^{\prec}, \omega^{\sim}.$$

Le modèle BFpref est capable d'exprimer les trois aspects de l'imperfection dans les préférences, et il distingue l'ambiguïté de la définition du terme « incomparabilité ». Le cas « non décisif » est directement représenté par un singleton alors que les informations manquantes (cas non décidé) sont représentées par un cas d'extrême imprécision - l'ignorance totale, qui est l'union de tous les singletons possibles sur le cadre du discernement.

En agrégation des préférences

À l'aide du modèle BFpref, nous avons proposé une stratégie d'agrégation des préférences en matière de preuve fondée sur la règle de combinaison de Dempster et une stratégie de décision de distance minimale. Nous avons également proposé une stratégie pour éviter le paradoxe de Condorcet ainsi qu'une méthode de parcours en profondeur (Depth First Search DFS en anglais) efficace pour la construction de graphes orientés acycliques. Comparant avec l'algorithme naïve, i.e. itérativement détecter circles et les éliminer, la DFS méthode performe mieux sur certain structures des graphes.

Lors de l'étape de décision de l'agrégation des préférences sur le modèle BFpref, nous avons signalé une faille dans les distances pour les objets évidentiels. Nous acceptons la propriété de « structurel fortement » car elles expliquent l'interaction entre les éléments focaux de deux fonctions de masse comparées. Cependant, aucune distance n'est capable de distinguer des singletons avec des poids différents. C'est défaut est obvious en mesurant la similarité entre les relations de préférence. Par exemple, avec ce défaut, la similarité entre « indifférence » et « préférence stricte » est équivalent que cela entre « préférence stricte » et « préférence stricte inversé ». Ce résultat est contre toute les distances pour les relations de préférence et les sens communs. (la première doit être inférieur que la deuxième.)

Pour résoudre ce problème, nous avons analysé les hypothèses sur lesquelles repose la distance de Jousselme, qui s'agit une distance populaire pour les fonctions de masse, et en avons supprimé une hypothèse non désirée, qui considère que les distances entre les différents singletons sont égales. En étendant la distance de Jousselme, nous avons proposé la distance WSD (Weighted singleton distance en anglais). La distance WSD tient compte des différences de similarités entre les différents singletons, et elle est de plus fortement structuré. Autrement dit, la mesure entre les unions des singletons considère aussi des similarités différentes entre les singletons. Nous avons également appliqué cette distance WSD dans l'étape de prise de décision sur le jeu de données SUSHI (Une enquête sur les préférences de sushi en Japon). Les résultats de la comparaison montrent que les décisions fondées sur le WSD sont plus raisonnables que celles fondées sur la distance de Jousselme.

En apprentissage des préférences

L'apprentissage sur les préférences évidentielle, plus précisément, le clustering des préférences issues du modèle BFpref, est aussi abordé dans le cadre de cette thèse. Nous représentons les profils des agents par des préférences incertaines via le modèle BFpref lorsque plusieurs sources de préférences conflictuelles sont prises en compte. Dans notre méthode, l'estimation de fonctions de masses d'un agent identique correspond à l'estimation du conflit entre les différentes sources d'informations. Plus précisément, les dissimilarités entre les différentes sources d'un agent identique sont considérées comme un degré d'ignorance de cet agent, et les relations représentées pour chaque source sont distribuées sur les éléments correspondants des fonctions de masse. Les similarités entre les agents sont fondées sur la somme de la distance de Jousselme sur toutes les couples d'objets étudiés, ce qui équivaut à la distance de Kendall dans le cas des préférences certaines.

Concernant la phase de classification, la méthode de k-means ne pourra pas être appliquée car seulement les distances entre couples sont fournies, et de ce fait le calcul du centroïde devient un problème NP-difficile. Par conséquent, les algorithmes indépendants du calcul de « centroïdes » sont les seuls à pouvoir être appliqués dans notre cas. Dans ce travail, la méthode Ek-NNclus (Evidential k-NN clustering) est appliquée. Ek-NNclus s'agit d'une méthode souple de clustering aussi basé sur théorie de fonction de croyance. L'évaluation du résultat du clustering réalisé est fondée sur le score de silhouette, en comparaison avec d'autres distances telles que Kendall et Euclidean.

En comparant avec une stratégie appliquant moyenne arithmétique des deux sources de préférences conflictuelles des agents, il est illustré que le modèle BFpref renvoie un meilleur résultat de clustering en termes de score de silhouette. Notre méthode est donc capable de détecter les communautés des agents même en présence de préférences incertaines.

Le clustering est effectuée jusqu'alors uniquement sur des préférences complètes. En effet, la distance pour les ordres évidentielles incomplets fondée sur la distance de Jousselme ne permet pas la classification de données incomplètes. Cela est causé par l'ignorance (partielle ou totale) exprimée dans les fonctions de masse. Prenons un exemple extrême, les valeurs manquantes représentées par l'ignorance totale. Mesurées par la distance de Jousselme, l'ignorance est considérée comme un cas d'opinion identique d'un agent à un autre. Donc les agents ayant majoritairement des préférences manquantes seront classifiés dans un même groupe. Ce problème existe avec toutes les mesures de distance et divergences appliquées dans la théorie des fonctions de croyance. Les valeurs manquantes sont estimées par des mesures de conflits, qui considèrent l'ignorance totale comme un élément neutre, i.e. le conflit entre l'ignorance totale et toutes les autres fonctions de masse est nul. Cette propriété a pour conséquence la convergence de toutes les classes vers l'ignorance. Ce qui fausse le processus de clustering.

EPartant de ce constat, nous avons proposé et prouvé un théorème d'impossibilité actant qu'il n'existe pas de règles de combinaison pertinente pour le calcul du centroïde parmi les objets évidentiels, qui respectent les propriétés de « cohérence de metric », « surjectivité de centroïde » et de « neutralité de l'ignorance ». Une conséquence de ce théorème est qu'il n'y pas d'algorithmes de classification non-supervisée fondés sur le principe du k-means adaptés aux fonctions de masse, i.e. tenant compte des imprécisions des objets. Par contre, nous avons proposé une idée préliminaire de clustering en abandonnant la propriété de « cohérence métrique » et en mesurant la similarité entre les fonctions de masse par les valeurs évidentiel. Ce travail à approfondir est laissé en perspective.

La perspective la plus proche est d'appliquer la notion d'ignorance dans la classification sur des données incomplètes et comparer avec des méthodes d'état-de-l'art. Le problématique de mesurer la similarité entre les fonctions de masse mentionnée auparavant doit être d'abord réglé. Dans nos plans, nous allons mesurer les similarités entre les objets évidentiels aussi par les fonctions de masse dans une nouvelle cadre de discernement, pour qu'elle puisse mesurer les différence dans états d'objets ainsi que dans leurs niveau d'incertitude et d'imprécision. Dans ce travail, l'estimation de valeur des fonctions de masse de mesure dans le nouvelle cadre de discernement. Les règles à respecter dans ce processus est aussi une aspect importante.

Perspectives

A part des contributions présentées ci-dessus, nous avons également proposé en thèse une idée préliminaire de regroupement des organes de preuve en abandonnant la propriété de « cohérence métrique ». Nous estimons que la similarité entre les organismes de preuve devrait également être mesurée par les valeurs de preuve dans un autre cadre de discernement du TBF. Cette idée devrait être explorée et vérifiée, ce qui reste dans nos perspectives à court terme.

De plus, le problème de la prise de décision est limité en mono-critère dans cette thèse. La prise de décision multicritères (MCDM) est un problème plus fréquemment rencontré. PROMETHEE est une célèbre méthode MCDM, et est également dans le cadre de la comparaison par paires entre les alternatives comme dans le modèle BFpref. Ainsi, l'extension de la BFpref sur PROMETHEE pour les problèmes MCDM serait une perspective prometteuse.

Contents

Part I

Introduction

Chapter 1

Introduction

> "To be, or not to be, that is the question."
>
> — William Shakespeare, Hamlet

In Shakespeare's play ***Hamlet***, the prince Hamlet faces two options: "to be, or not to be". His selection between the two options is a simple expression of his preference. His mechanism of analyzing the benefits and loss in taking the choice is studied in the decision making theory.

From the selection between objects, more information other than preference opinion on objects to be selected may also be potentially implied. Here is an example in presidency election. In the USA presidency election in 2016, in addition to the result that Donald Trump was elected the 45th U.S. president, there are also research [FSW+17] showing that the choice (which can be regarded as a sort of preference) of voters may correlate with their class, race, gender or other elements. The preferences of a person may relate to his/her potential properties, such as characteristics, education back ground, communities, hobbies, ***etc***. Moreover, not only the preference itself, the certain of preference may also reflect such third-part information. In psychology, the uncertainty of choice is used in the determination of personality types, such as the DiSC® personality test applications [Sug09].

All these examples fall on the management of preference information. In this thesis, we study the modeling and management of imperfect preference information, with uncertainty, imprecision and incompleteness.

1.1 Brief overview on preferences

Preferences play an important role in the activities of human beings, especially in psychology, economics, sociology and philosophy. For instance, in democratic politics, delegates are elected by voting and such systems are essentially procedures of aggregation of preferences. In economics, preference relation is usually used for the comparison between alternatives, applied in decision making procedure. In sports matches, the ranking result of athletes is a format of preferences. In computer science and technology, rankings of objects, such as searching results, are also types of preferences.

The study on preference has a long history. It has been under the high light of researchers for centuries , especially in voting systems. Here we list some historic timestamps of some important initiative research on preferences and more details will be discussed in Chapter II. Early to medieval period in 1299, Ramon Llull, considered as the earliest founding father of voting theory and social choice theory, invented the voting method named "***Ars generalis***" (meaning general method), also known as "Copeland's method" based on binary combination of simple alternatives [Col13]. In 1435, Nicholas of Cusa is believed to be the first person inventing Borda count method (named after the French mathematician Jean-Charles de Borda) for preference aggregation [HP08]. In 18th century, Marquis de Condorcet's studied the cyclic cases in preference aggregation, namely Condorcet's Paradox [dCmdC85]. In 1951, Nobel laureate Kenneth Arrow found an impossibility theorem, namely Arrow's impossibility theorem, pointing a vital problematic in the study of preferences.

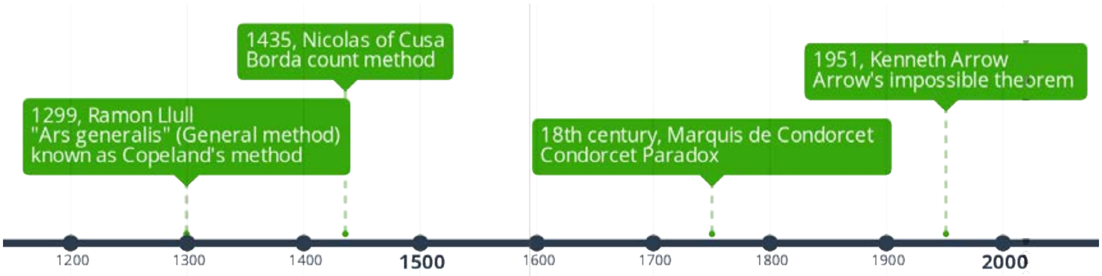


Figure 1.1 – Important study on preference in history.

Nowadays, the study on the preference has pivoted to the model of unconventional preferences and novel applications on preference management. Fuzzy set theory [Zad65] has been introduced to model the uncertain preference and many methods on multi-criteria decision making have been proposed such as AHP (Analytic Hierarchy Process) by Saaty in 1970s [Saa02], Electre (ELimination Et Choix Traduisant la REalité) by Roy in 1968 [Roy68], PROMETHEE by Brans and Vincke in 1985 [BV85] and other methods. Today, many research works about preference focus on the application of recommendation systems. Netflix Prize [BL$^+$07], a well awarded competition for recommendation systems, has set fire on the blossom of the applications on preferences.

Before the introduction of more detailed technique issues, we unify the terminology of this thesis first. In a simple piece of preference information, the compared objects are named as "alternatives". [1] The entity who express the preferences are named as "agent"[2].

[1]The term "items" is usually applied in recommendation systems and "candidate" in electoral systems.

[2]The term "user" or "individual" is usually applied in recommendation systems and "voter" in

The preference management mainly consists of two aspects: preference aggregation, which is also a sub-domain of decision making theory, and preference learning, which is a sub-domain of machine learning.

1.1.1 Preference aggregation

Most of the conventional study on preference were focused on problems about preference aggregation methods. which is also named social choice theory. In this process, multiple preferences are aggregated into one, usually applied for decision making. Here is a simple example of preference aggregation with a problem of paradox:

Example 1. Which activity to practice?

> Alice, Bob, and Charlie decide to organize an activity for their Saturday afternoon. They vote among three options: practicing Archery (a), playing Basketball (b), and watching a movie in Cinema (c). By $a \succ b$ denote the proposition that outcom a is preferred to outcome b. The opinion of each one is given as:

$$\begin{aligned} Alice: &\quad a \succ b \succ c \\ Bob: &\quad b \succ c \succ a \\ Charlie: &\quad c \succ a \succ b \end{aligned} \tag{1.1}$$

One straightforward approach would be to pick the activity with the largest number of votes (namely plurality method). However, by applying such method, we encounter a cyclic condition, making it impossible to make the final decision. This paradox is named as Condorcet's paradox.

To conclude, social choice theory deals with the problems on the aggregation of preferences and decision making. In this thesis, we focus only on mono-criterion decision making problem, where every compared item is represented by only one single criterion. Concerning this topic, in Chapter 4, state-of-the-art methods of mono-criterion decision making are introduced. Besides, aggregation on preferences with uncertainty is discussed in Chapters 5 and 6.

1.1.2 Preference learning

Nevertheless, the study on preferences is not limited to social choice theory. With the trendy development of artificial intelligence (AI), reasoning with preference is considered as a particularly promising research direction for the AI community [Doy04]. Preference learning is a relatively large topic. Roughly speaking, it is about inducing predictive preference models from empirical data. Some definitions on preference

electoral systems.

learning are larger, including also the application of preferences in machine learning algorithms [FH11]. In this thesis, we focus mainly on the former one. Here is a simple example on preference learning.

Example 2. Chopsticks or fork and knife

> Table manner and habits are different between China and France. In China, people prefer chopsticks to fork and knife as eating utensil while in France, fork and knife are always preferred. David is a server in a restaurant offering both kinds of utensils. By learning this knowledge on eating utensil preference, when a Chinese client comes, David elicit that chopsticks are more probably preferred and he would prepare chopsticks as default utensil.

Similar to this example, preference learning techniques are widely used from searching engine services to recommendation systems building. To find groups among agents upon their preferences is a fundamental step in preference elicitation, also named as preference clustering. More discussion on preference learning is given in Chapter 4. Besides, in Chapter 7, we study more cases of clustering on unconventional preferences.

In the two examples given above, preferences are expressed in crystal clear ways, meaning that all preference relations between alternatives are certain. However, such conditions are no longer enough to solve problems encountered nowadays, coming up with new challenges.

1.2 New challenges: Imperfectness of agent preference data

Initiated by the development of digital technology, especially the boom of the Internet, more possibilities of collecting and mining the preference data become possible. Such development has proposed abundant of novel challenges and issues on the study of preferences, varying from theoretic to application levels. The issues include but are not limited to the imperfectness of the preference data and the corresponding applications.

In the scope of this thesis, the imperfectness on preference data consists of mainly three aspects:

1. Preference with uncertainty;

2. Preference with imprecision;

3. Incomplete preference.

The uncertainty in preferences refers to epistemic situations on the knowledge of preference information where the preference relations can not be exactly described. The imprecision in preferences refers to the cases where multiple preference relations are

possible, usually caused by absence of knowledge or implicity on preference information. The incompleteness in preference refer to the cases that preference relation are not observed on all elements in a alternative space.
In this thesis, the term "evidential" is used, referring to circumstances with both "imprecision" and "uncertainty". For better understanding, some corresponding examples are given as follows.

1.2.1 Preferences with uncertainty

The uncertainty in preferences can be interpreted in two ways: uncertainty expressed directly from agents , which is in terms of cognition, or uncertainty in the preference data, which is in terms of statistic. Here are two examples respectively demonstrate these two ways of interpretation.

Example 3. Ramboutan and apple:
Uncertain preference caused by lack of knowledge.

> Alice is asked to express her preference between ***ramboutan*** and ***apple***. Alice has never tasted ramboutan but she infers the taste of ramboutan by the similarity in term of shape between ramboutan and litchi. As Alice knows she prefers litchi to apple, she gives an uncertain opinion that she prefers ramboutan to apple.

Figure 1.2 – Photo of litchis and ramboutans.

The three pieces of fruits on the left with bold peels are litchis. The other three with long hair are ramboutans.

Example 4. SUSHI preference data set: conflicting sources of preference data

> In the SUSHI preference data set [Kam03a], every respondent is asked to express his/her preference over different sushis by both ranking and scoring. Between two types of sushi ***Maguro*** and ***Ebi***, Bob expresses ***Maguro*** outranks ***Ebi*** in ranking but gives ***Maguro*** 3 points and ***Ebi*** 4 points, indicating that ***Maguro*** $\prec$ ***Ebi***. In this way, Bob's preference between ***Maguro*** and ***Ebi*** is uncertain

The reasoning methods on uncertain knowledge are various. In Section 2.2 of the Chapter 2, several popular reasoning methods on uncertain preference are introduced.

(a) Sushi Ebi

(b) Sushi Maguro

Figure 1.3 – Two types of sushi Ebi and Maguro.

1.2.2 Preferences with imprecision

The imprecision in preference implies that more than one preference relation may exist between a pair of alternatives under comparison. Here we give examples of imprecise preference inferred from natural language.

Example 5. Imprecision in preferences

> The sentence "A Peugeot 208 car is not worse than a Renault Clio" implies that ***Peugeot*** is preferred or indifferent to ***Renault***. The sentence "Marvel Comics and DC Comics are not in the same class" indicates only that ***Marvel*** and ***DC*** are not indifferent, but both ***Marvel*** is preferred to ***DC*** or ***DC*** is preferred to ***Marvel*** are possible.

1.2.3 Incomplete preferences

Sometimes, agents may not give their preference opinion on the whole set of alternatives. The incompleteness in preferences is caused by missing information. An example on real world data is given here:

Example 6. Missing information in Sushi preference dataset

> Again, in Sushi preference dataset, every respondent is asked to give their preferences on 10 random sushis out of 100. Thus, the missing information on the left 90 sushis causes incompleteness in preference.

However, some works consider that "incompleteness" is caused by the "incomparability" relation [RS93]. Such controversy is also due to the different interpretation on "incomparability". Some works regard "incomparability" as the absence of knowledge. The clarification on these notions are also in the scope of this thesis with an unified framework proposed, based on the theory of belief functions. More detailed discussions are given in Chapters 3, 5 and 6.

Targeted on these issues on the imperfectness of preference information, the objectives of this thesis are as follows.

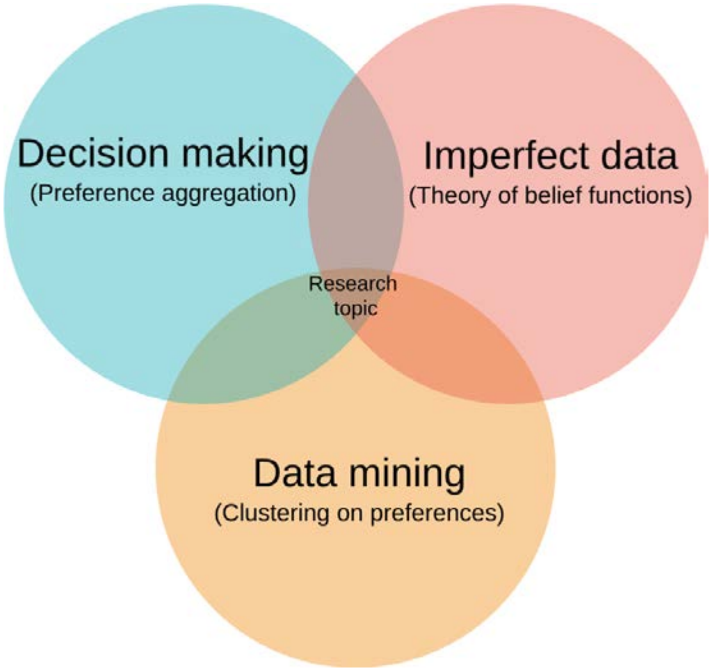


Figure 1.4 – Research topic of this thesis.

1.3 Objectives and contributions

The goal of this thesis is to improve preference models for imperfect preference and explore management problems on such preferences. To deal with problems of uncertainty and imprecision, we applied the theory of belief functions.

The theory of belief functions is a formal framework for representing and reasoning with uncertain and imprecise information by extending both the set-membership approach and probability theory. More detailed introduction is given in Chapter 2. The research topics of this thesis are concluded by Figure 1.4. The topics mainly combines the modeling on imperfect preference data (with the theory of belief functions), decision making from imperfect data (preference aggregation) and data mining on preference (clustering on preference data)

A global view of problematic in this thesis is illustrated in Figure 1.5. The blocks in grey marks the contributions, which mainly concern the domains of modeling on imperfect preferences as well as management (including aggregation and clustering) of such preferences,.

Agents may express their preferences with imperfectness, caused by conflicting sources, uncertainty or the ambiguity of expression. Several contributions are made to the field of imperfect preference modeling and management. In modeling part, we propose the BFpref model, an evidential model for imperfect preference based on the theory of belief functions. Based on BFpref model, we studied the management prob-

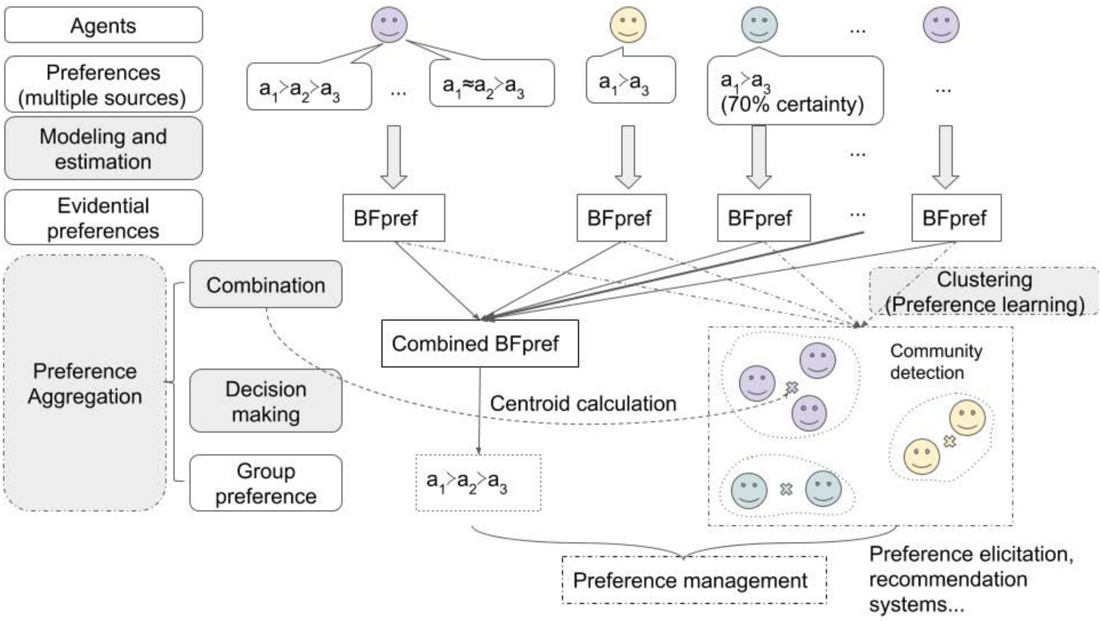


Figure 1.5 – Problems and processes concerned in the thesis

lems from aggregation to clustering of imperfect preferences.

Modeling of imperfect preference

We firstly clarified some ambiguity at the definition level on different types of preference relation, especially the definition of "incomparability". Besides, raw preference data may be imperfect from agents level: uncertain, conflicting, and incomplete preferences are possible in agents' expression. To model such preferences, we propose a general framework expressing the "uncertainty", "imprecision" and "incompleteness" in preference with the help of the theory of belief functions (TBF), named as BFpref model. Basic Belief Assignment (BBA) is used to express the uncertainty and imprecision level on preference relations. More introduction on TBF is given in Section 2.3 of Chapter 2.

Aggregation problems on imperfect preferences

This contribution mainly consists of two aspects. Firstly, we proposed and compared different aggregation strategies, based on different combination rules, and decision-making methods for BFpref model, as the theory of belief functions (TBF) has been proved to be a practical mathematical tool for data fusion. However, we find that the state-of-the-art distance measures in TBF are not adaptable for decision in preference relations. An important one is that all singletons are equally measured, making the distance between "indifference" and "strict preference" equal to the distance between two "strict preferences inverse". To solve this problem, a new metric named Weighted

Singleton Distance (WSD) is proposed, enabling to measure the similarity between two BBAs in the same framework with different weights on singletons. In the solution of avoidance of Condorcet's paradox (introduced in Section 4.2 of Chapter 4), we also proposed a faster algorithm for Directed Acyclic Graph (DAG) building based on Depth First Search (DFS), which improves the efficiency of cycle detection and elimination in evidential preference structures.

Clustering on evidential preferences

In this part, BFpref as well as different distances are applied in the application of community detection. Community detection is a popular topic in network science field. In social network analysis, preference is often applied as an attribute for individuals' representation. In some cases, uncertain and imprecise preferences may appear. Moreover, conflicting preferences can arise from multiple sources. Based on BFpref model, the clustering quality in case of perfect preferences as well as imperfect ones based on weak orders (orders that are complete, reflexive and transitive) are studied. Limited by several properties, the clustering process is only executed on complete preference orders. Some necessary properties for incomplete preference orders are also discussed and are in the scope of our future work.

Table 1.1 shows the topics on evidential preference management, including preference aggregation and clustering. The preferences are categorized into two parts: complete and incomplete. In this thesis, only aggregation on both categories of preferences are studied. The clustering on preference is limited on complete preferences while only discussion is given for incomplete preferences.

Table 1.1 – Preference management concerned in this thesis

	group decision making	clustering
complete	$\surd$	$\surd$
incomplete	$\surd$	discussion

1.4 Structure of the thesis

This thesis is organized in three parts, including eight chapters. In addition to the first chapter made up of the introduction, Part II and III respectively consist of four and three chapters.

In Part II, the state of the art methods on preference modeling, preference aggregation as well as preference learning on mono-criterion preferences are presented.
In Chapter 2, the basic models of preferences are firstly introduced. A brief introduction on the theory of belief functions is also given.
In Chapter 3, different similarity measure methods are introduced. These measures includes distances for preference relations, distances and correlations for preferences

structures (orders) as well as distance, conflict and divergence measure methods in TBF.
Chapter 4 concerns the management of preferences, including preference aggregation and preference learning. In the first aspect, different preference aggregation methods as well as problematic issues are introduced. Besides, the distances play important roles in both aggregation and preference learning. Thus, distances used in preference and in the theory of belief functions as well as preference learning are also introduced in this chapter. In the second aspect, a brief introduction on preference learning is presented, along with applications on preference clustering. Such techniques are often applied in preference based community detection and recommendation systems.

The contributions of the thesis are introduced in Part III.
In Chapter 5, some ambiguities in the definitions of preference relations, especially the "incomparability", are discussed and clarified. Based on this clarification, an evidential preference model based on the theory of belief functions (namely BFpref) is proposed, enabling the expression of preferences with uncertainty and imprecision. Based on BF-pref model, an aggregation method for evidential preferences is proposed and compared with other conventional procedures. A Condorcet's paradox avoidance method is also introduced, with a faster algorithm for DAG building proposed.
In Chapter 6, we focus on a flaw in the similarity measure, introduced above, on BBAs and proposed a new metric named WSD with the issue solved. WSD is applied in the aggregation of imperfect preferences and the comparison with other methods show that WSD is more reasonable.
In Chapter 7, the clustering applications on preferences expressed by BFpref model is studied, with a new evidential preference reasoning method proposed. Some comparisons between different metrics are also illustrated.

Finally, the last chapter gives conclusions and perspectives on our work.

During the work of this thesis, we have published the following papers.

1.5 Publication list

International conferences

- **Zhang Y.**, Bouadi T., Martin A. Preference fusion and Condorcet's Paradox under uncertainty. ***20th International Conference on Information Fusion (FUSION), Xi'an, 2017, pp. 1-8***
- **Zhang Y.**, Bouadi T., Martin A. A clustering model for uncertain preferences based on belief functions ***International Conference on Big Data Analytics and Knowledge Discovery. Springer, Cham, 2018: 111-125.***
- **Zhang Y.**, Bouadi T., Martin A. An empirical study to determine the optimal k in Ek-NNclus method. ***International Conference on Belief Functions. Springer, Cham, 2018.***

International journals

- **Zhang Y.***, Bouadi T., Martin A. A distance for uncertain and imprecise preferences based on the theory of belief functions, applied for group decision making. Under review of ***Journal of Decision Support Systems***
- Ma Z., Liu Z., **Zhang Y.***, Luo C., Song L. Credal Transfer Learning with Multi-estimation for Missing Data. ***IEEE Access (2020)***
- Zhang Z.*, Tian H., **Zhang Y.**, Ma Z., He J. Incomplete Pattern Credal Classification with Optimize and Adaptive Multi-Estimation in Different Domains. Under review of ***IEEE Transactions on Fuzzy Systems***

Part II

Background

Chapter 2

Basic concepts

To simplify the expression of the work in this thesis, we introduce basic concepts and models around the preference paradigm in this chapter. Besides, a brief introduction on the theory of belief functions is given as well. The chapter is structured as follows. Concepts of preference modeling are introduced in Section 2.1, followed by an introduction of different modeling on unconventional preferences in Section 2.2. An introduction on the theory of belief functions (TBF) in Section 2.3, with basic definitions and several commonly used combination rules given.

2.1 Preference Modeling

The representation of preferences has been studied in various domains such as decision theory [ÖTV05], artificial intelligence [WD91], economics and sociology [Arr59]. Preferences are essential to efficiently express user's needs or wishes in decision support systems such as recommendation systems and other preference-aware interactive systems that need to elicit and satisfy user preferences. However, preference modeling and preference elicitation are not easy tasks, because human beings tend to express their opinions in natural language rather than in the form of preference relations. Preferences are also widely used in collective decision making and social choice theory, where the group's choice is made by aggregating individual preferences.

2.1.1 Binary relations

The preference are mostly modeled in forms of binary relation between two alternatives.

Definition 2.1.1. Binary Relation: A binary relation R on a set $\mathcal{A}$ is a subset of the cartesian product $\mathcal{A} \times \mathcal{A}$.

In preferences, $\mathcal{A}$ denotes the set of alternatives and R the preference relation. The product of binary relations can be regarded as transitive between multiple binary relations in a sequence, defined by the following definition.

Definition 2.1.2. Product of binary relations: Given two binary relations R_1 and R_2, three alternatives $a_1, a_2, \in \mathcal{A}$: $a_1 R_1 \cdot R_2 a_2 \Leftrightarrow \exists a_i \in \mathcal{A} : a_1 R_1 a_i$ and $a_i R_2 a_2$

2.1.2 Properties of binary relations

Before introducing different preference structures, we collectively cite different possible properties for a binary relation.

A relation R is called:

reflexive, if	$aRa, \forall a \in A$
irreflexive, if	$a\neg Ra, \forall a \in A$
symmetric, if	$aRb \rightarrow bRa, \forall a, b \in A$
antisymmetric, if	$(aRb, bRa) \rightarrow a = b, \forall a, b \in A$
complete, if	$(aRb$ or $bRa), \forall a \neq b \in A$
strongly complete, if	$(aRb$ or $bRa), \forall a, b \in A$
transitive, if	$(aRb, bRc) \rightarrow aRc, \forall a, b, c \in A$
negatively transitive, if	$a\neg Rb, b\neg Rc \rightarrow a\neg Rc, \forall a, b, c \in A$
semi-transitive, if	$(aRb, bRc) \rightarrow (aRd$ or $dRc), \forall a, b, c, d \in A$
Ferrers relation, if	$(aRb, cRd) \rightarrow (aRd$ or $cRb), \forall a, b, c, d \in A$

Several preference relations are defined depending on the properties they satisfy, introduced as follows.

2.1.3 Preference relations

Based on the definition of a binary relation, between any couple of alternatives a_i, a_j (without order), only 4 relations possibly exist $\{\succ, \succeq, \approx, \sim\}$:

- $\succ$: Strict preference. $a_i \succ a_j$ states that "a_i is preferred to a_j" Between a_i, a_j both $a_i \succ a_j$ and $a_j \succ a_i$ are possible. For ease of readability, we use $a_i \prec a_j$ instead of $a_j \succ a_i$ in the following parts
- $\approx$: Indifference. $a_i \approx a_j$ states that "a_i is indifferent to a_j"
- $\succeq$: Weak preference. $a \succeq b$ states that "a_i is preferred or indifferent to a_j"
- $\sim$: Incomparability. $a_i \sim a_j$ states that "a_i is incomparable to a_j"

Weak preference is the union of ***strict preference*** and ***indifference***. Preferences on a set of alternatives are usually represented by a ***preference structure***, defined as follows.

Definition 2.1.3. Preference structure: A preference structure on $\mathcal{A}$ is a reflexive binary relation R on $\mathcal{A}$. As the *complete* property is implied by the existence of the incomparability relation, we categorize different preference structures by incomparability. *i.e.* structures without incomparability are complete while structures with incomparability are considered as incomplete.

These relations can be defined on binary relation R as:

$$\begin{aligned} a_i \succ a_j &\Leftrightarrow a_i R a_j \text{ and } a_j \neg R a_i \\ a_i \approx a_j &\Leftrightarrow a_i R a_j \text{ and } a_j R a_i \end{aligned}$$

With the definitions of $\langle \succ \rangle$ and $\langle \approx \rangle$, we have two other relations:

$$\begin{aligned} a_i \succeq a_j &\Leftrightarrow a_i \succ a_j \text{ or } a_i \approx a_j \\ a_i \sim a_j &\Leftrightarrow a_i \neg \succ a_j \text{ and } a_i \neg \prec a_j \text{ and } a_i \neg \approx a_j \end{aligned}$$

The "weak preference" is the union of strict preference and indifference. Thus, between two alternatives a_i and a_j, the four relations $\{\succ, \prec, \approx, \sim\}$ are exclusive and exhaustive[1]. We propose the illustration given in Figure 2.1. In this figure, the round on the left represent the set of relation $a_i R a_j$ and the round on the right the set of relation $a_j R a_i$. The overlapped part, showing the intersection of the two sets, represent the set of relation $a_i \approx a_j$. According to the definition, the incomparability relation $a_i \sim a_j$ is the negation of all the three relations above, thus represented by masked part in orange.

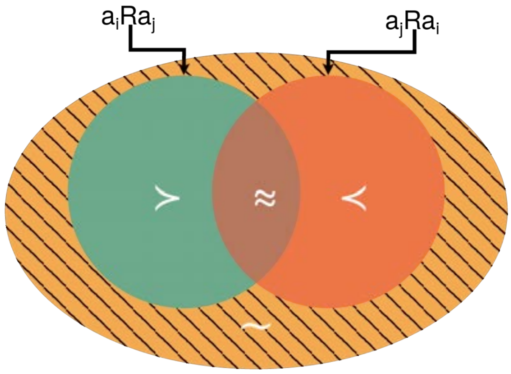


Figure 2.1 – Venn diagram of preference relations

From the definitions above, the following properties are guaranteed:

- $\langle \succ \rangle$ is transitive and anti-symmetric;
- $\langle \approx \rangle$ is transitive, symmetric and reflexive

With the relations "Strict preference","Indifference" and "Incomparability", and R a binary relation on the alternative set A, we can define different preference orders:

Definition 2.1.4. Total order: R is a total order if and only if $R \in \{\succ, \prec\}$

A total order is also called "linear" order.

Definition 2.1.5. Weak order: R is a weak order if and only if $R \in \{\succ, \prec, \approx\}$

[1] As $a_i \succ a_j$ is equivalent to $a_j \prec a_i$, to avoid repetitive comparisons between two alternatives, we assume that $i < j$ in this article.

Definition 2.1.6. Partial order: R is a partial order if and only if $R \in \{\succ, \prec, \sim\}$

In many works, ***weak orders*** are also referred to as ***partial rankings***. In our work, we accept the terms "weak order" and "partial order".

Definition 2.1.7. Quasi (pre)order: R is a quasi-(pre)order if and only if $R \in \{\succ, \prec, \approx, \sim\}$

Examples

For better understanding, we give some examples from real life of each type of orders.

Example 7. Total order: Sprint race

> In a sprint race in track and field competitions, final ranking of competitors is a total order.

Example 8. Weak order: Exam for students

> In an exam, if we rank all students by their notes, this ranking is a weak order. Indifference relations are caused by several students having same notes.

The example on partial order can be ambiguous, caused by the definition of "incomparability". Different circumstances of "incomparability" may be related to different interpretations. As this is still an open issue, we give the corresponding example in the contribution part in Chapter 5.

To distinguish conventional preference and in-conventional ones introduced later, we use the term "***crisp***" and "***soft***". For crisp preference, every relation is binary. For soft preference, ***uncertainty*** may exist. Thus binary representation is no longer enough. Besides, imperfectness such as ***imprecision*** and ***ignorance*** may also exist. In next section, we introduce different theories for soft preference modeling.

2.2 Uncertain preferences

Almost all soft methods can be applied in uncertain preference modeling, such as probability theory, fuzzy set theory, rough set theory, ***etc***. In this section, we introduce several classical models for soft preference modeling.

2.2.1 Fuzzy preference

Confronting preferences with imprecision, the theory of fuzzy set [Zad65] has always been a powerful tool. The most accepted model was firstly proposed by Blin [Bli74], and then developed by others on different applications, notably in group decision making [BSS78, Tan84].

Fuzzy preference modeling is based on fuzzy membership. We borrow the definitions in [Bli74]. A fuzzy binary relation on an alternative set $\mathcal{A}$ is determined by a fuzzy set on the product set $\mathcal{A} \times \mathcal{A}$, that is, by a membership function (MF) $\mu_R : \mathcal{A} \times \mathcal{A} \to [0,1]$ over the set $\mathcal{A} \times \mathcal{A}$. Considering the case where $\mathcal{A}$ is finite, $\mathcal{A} = \{a_1, \ldots, a_N\}$, we can define an $N \times N$ matrix $\mathbf{R}$ whose (i,j) element r_{ij} is given by:

$$r_{ij} = \mu_R(a_i, a_j), i, j = 1, \ldots, N \tag{2.1}$$

from the membership function μ_R. Conversely an $n \times n$ matrix with elements in $[0,1]$ defines a fuzzy binary relation on $\mathcal{A}$. Thus, we have $0 \leqslant r_{ij} \leqslant 1$.
When R is a preference relation, the element r_{ij} represents the degree of preference of alternative a_i to alternative a_j. There are two ways of interpretation on the values of r_{ij}: One is that $r_{ij} > 0.5$ represents an uncertain preference of a_i over a_j while a_i is definite (meaning without uncertainty) preferred to a_j if $r_{ij} = 1$. The other interpretation is that $r_{ij} > 0.5$ represents a preference of a_i to a_j with some intensity level, while $r_{ij} = 1$ represents the highest intensity level.

Most of works on fuzzy preferences do not distinguish the concepts of “imprecision” and “intensity”. They believe that the two interpretations are rather homogeneous because a certain opinion of “prefer” often corresponds to an intense preference.

In fuzzy set theory, some assumptions are accepted.

- $\mathbf{R}$ is reciprocal. It can be additive:

  $$r_{ij} + r_{ji} = 1, \tag{2.2}$$

  or multiplicative [HVCHA07, Xu15]:

  $$r_{ij} \cdot r_{ji} = 1 \tag{2.3}$$

- Intensity is transitive: if x_i is preferred to x_j and x_j is preferred to x_k, then x_i should be preferred to x_k with at least the same intensity. Formally:

  $$\begin{aligned} &r_{ij} \geqslant 0.5, r_{jk} \geqslant 0.5 \\ \Rightarrow\ &r_{ik} \geqslant min(r_{ij}, r_{jk}), \forall i, j, k \in 1, \ldots, N \end{aligned} \tag{2.4}$$

  This type of transitivity is also called ***moderate stochastic transitivity*** in the probabilistic choice theory [Luc12].

Some extension versions of fuzzy preferences are also proposed. For example, the authors of [Xu07] introduced intuitionistic fuzzy preference relations (IFPR) based on intuitionistic fuzzy set [Ata99] by introducing both membership function and non-membership function. Traditionally, fuzzy preference models are used in the context of preference aggregation applications in decision making theory. The corresponding methods are introduced in Chapter 4.

Fuzzy set theory is popular in reasoning uncertain information. Other soft method theories have also been applied in preference modeling.

2.2.2 Preference by probability: Bayesian model

Instead of reasoning, Baysian model is a popular probabilistic model usually applied in preference elicitation. The model is defined as follows: For agent n, given a utility function $g(a_i)$ on alternative a_i, on pairwise preference a_i, a_j, a common model for observations $a_i \succ a_j$ is the probit likelihood [GSB10] defined as:

$$p(a_i \succ a_j) = \Phi(g(i) - g(j)) \tag{2.5}$$

where Φ denotes the standard Gaussian cumulative distribution function.
Another popular option is the logit likelihood [KP13] defined as

$$p(a_i \succ a_j) = \frac{1}{e^{-(g(i)-g(j))}} \tag{2.6}$$

Bayesian model is often applied in Plackett-Luce ranking model [Luc59, Pla75]. The model is defined as follows:
Given rankings σ on set $\mathcal{A} = \{a_1, \ldots, a_N\}$ Under Luce's axiom, the probability of selecting alternative a_j from $\mathcal{A}$ is given by

$$P(a_j|\mathcal{A}) = \frac{\alpha(a_j)}{\sum_{a_i \in \mathcal{A}} \alpha(a_i)} \tag{2.7}$$

where $\alpha(a_i)$ denotes the worth of alternative a_i.

This model is usually applied in social choice problems (which is introduced in Chapter 4)

Similar to Fuzzy set theory, Bayesian model are not able to express ignorance in preference information. In the cases with ignorance, the theory of belief functions (TBF) becomes more useful. TBF is a powerful mathematical tool for modeling uncertain data, ***i.e.*** data with both ignorance and imprecision.

2.3 Introduction to theory of belief functions (TBF)

The theory of belief functions (TBF) (also referred to as Dempster-Shafer or Evidence Theory) is a mathematical theory that generalizes the theory of probabilities by giving up the additivity constraint. It was firstly introduced by Dempster [Dem67] in the context of statistical inference as a general model of uncertainties. Afterwards, Shafer [Sha76] formalized it as a theory of evidence. In 1980's and 1990's, Smets popularized and developed this theory by proposing Transferable Belief Model (TBM) [Sme90] and it has been applied widely in various domains such as information fusion, classification, reliability and risk analysis, ***etc.*** By extending probabilistic and set-valued representations, it allows to represent degrees of belief and incomplete information in a unified frame.
Unlike the probability theory, which is unable to distinguish equally probable events from the case of ignorance, in the theory of belief functions, both imprecision and ignorance are modeled.
TBF has been proved to be a powerful tool in applications of data fusion [Sme00, EMS04, KKKR13]. Generally, data fusion consists of four steps:

1. Modeling
2. Estimation
3. Combination
4. Decision

In this section, we introduce basic definitions and tools in the theory of belief functions accompanied with a tour-guide for data fusion.

Definition 2.3.1. Discernment frame: The discernment frame Ω is a finite set of disjoint elements, defining the domain of references, formally:

$$\Omega = \{\omega_1, \dots, \omega_n\} \tag{2.8}$$

where ω_i are exclusive and exhaustive.

The definition of hypotheses ω are related to the given problem. The belief functions are defined on the power set $2^{\Omega} = \{X : X \subseteq \Omega\}$. Confronting a data fusion problem, the modeling step is fundamentally on the definition of a proper frame of discernment.

Definition 2.3.2. Basic Belief Assignment[2]: Function $m : 2^{\Omega} \to [0, 1]$ is called *Basic Belief Assignment (BBA)* on 2^{Ω}, such that

$$\sum_{X \subseteq \Omega} m(X) = 1 \tag{2.9}$$

A BBA is called ***normalized*** if $m(\emptyset) = 0$.

Definition 2.3.3. Categorical BBA: A *categorical BBA* is a normalized BBA which has a unique focal element X^*, formally:

$$m(X) = \begin{cases} 1, & \text{if } X = X^* \subset \Omega \\ 0, & \textit{otherwise} \end{cases} \tag{2.10}$$

Categorical BBAs are specific cases of simple BBA. Thus a categorical BBA on element $X, X \subset \Omega$ is denoted as X^0.

In a BBA, the values on union elements representing the imprecision are also interpreted as the ignorance of the knowledge, formally, $m(X) \neq 0, |X| > 1$. With the imprecision, when $m(\Omega) < 1$ we call m represent ***partial ignorance*** on X. Otherwise, if $m(\Omega) = 1$, m is called “vacuous”, representing ***total ignorance***.

Definition 2.3.4. Vacuous BBA: A *vacuous BBA* is a particular categorical BBA focused on Ω, formally:

$$m(X) = \begin{cases} 1, & \text{if } X = \Omega \\ 0, & \textit{otherwise} \end{cases} \tag{2.11}$$

[2]Basic Belief Assignment is also called *mass function*

Vacuous BBA are denoted as $m_?$ in this thesis.

Definition 2.3.5. Focal element (focal set): Every element $X \subseteq \Omega$ such that $m(X) > 0$ is called focal set, and focal element if $X \in \Omega$.

Definition 2.3.6. Simple BBA: m is a simple BBA if it has the following form:

$$\begin{aligned} m(X) &= 1 - w(X) \\ m(\Omega) &= w(X) \end{aligned} \tag{2.12}$$

where $w(X)$ denote the weight function (Equation (2.19)) on X. for some $X \subset \Omega, X \neq \emptyset$ and $w(X) \in [0,1]$. For convenience, such BBA is denoted by $X^{w(X)}$.

Thus, a BBA is capable to express both uncertainty and imprecision. When the focal element has value less than 1, the uncertainty level is expressed. While the focal element is on the union of singletons, imprecision is expressed.

Definition 2.3.7. Dogmatic BBA: A dogmatic BBA is a BBA where Ω is not a focal element. Formally:

$$m(\Omega) = 0 \tag{2.13}$$

Definition 2.3.8. Bayesian BBA: A *Bayesian BBA* is a BBA which all focal elements are elementary hypotheses, *i.e.* all focal elements are singletons. Formally:

$$m(X) = \begin{cases} \in [0,1], & \text{if } |X| = 1 \\ 0, & otherwise \end{cases} \tag{2.14}$$

A Bayesian BBA is a probability distribution over frame Ω. Furthermore, if a Bayesian BBA is categorical, it describes that there is no uncertainty. Thus the state of the concerned variable is certain and precise.

In data fusion procedures, the value assignment on BBAs is the estimation step.

Definition 2.3.9. Belief function: If the evidence tells us that the truth is in Y, and $Y \subseteq X$, we say that the evidence supports X.
Given a normalized BBA m, the probability that evidence supports X is:

$$Bel(Y) = \sum_{X \subseteq Y} m(X) \tag{2.15}$$

The value of $Bel(X)$ is called the degree of belief in X, and the function is called a belief function.

Definition 2.3.10. Plausibility function: If the evidence does not support $\bar{X}$, it is consistent with X.
For a normalized BBA m (*i.e.* $m(\emptyset) = 0$), the probability that the evidence is consistent with X is

$$Pl(X) = \sum_{X \cap Y \neq \emptyset} m(Y) \tag{2.16}$$

$$= 1 - Bel(\bar{X}) \tag{2.17}$$

The value of $Pl(X)$ is called the plausibility of X and the function plausibility.

Definition 2.3.11. Commonality function: Commonality function $Q : 2^{\Omega} \to [0, 1]$ is another equivalent representation of a belief function. It is defined as:

$$Q(X) = \sum_{Y \supseteq X} m(Y), \forall X \subseteq \Omega \tag{2.18}$$

With commonality function Q, the weight function from a BBA is defined.

Definition 2.3.12. Weight function: The weight function on element $X \subset \Omega$ is calculated by :

$$w(X) = \prod_{Y \supseteq X} Q(Y)^{(-1)^{|Y|-|X|+1}} \tag{2.19}$$

In the next part, we introduce TBF tools for combination and decision steps in data fusion.

2.3.1 Combination rules in TBF

As aforementioned, data fusion is an important application of the theory of belief functions. We borrow the definition of data fusion in [HAMA16]:

> Data fusion is the process of integrating multiple data sources to produce more consistent, accurate, and useful information than that provided by any individual data source.

In the theory of belief functions, the process of data fusion is usually related to the combination of multiple BBAs defined on the same frame of discernment, representing multiple sources on identical variables.
Many combination rules have been proposed with the account of various properties of information sources. Here among which an important one is introduced, namely ***cognitive independence***[Sme93].

Definition 2.3.13. Cognitive independence: Sources are considered as *cognitively independent* if any source has no communication with the others.

Cognitive independence is different from ***conditional independence***. The former one is defined from a point of view of the intrinsic property from the sources, while the latter one is defined from a statistical view.

For cognitive independent sources, when all sources are reliable, Dempster's combination rule and conjunctive combination rule are often applied.
In the discernment frame Ω, given two cognitively independent and reliable sources s_1 and s_2, m_1 and m_2 are two BBAs of s_2 and s_2 on Ω, the combination rules are defined as follows.

Conjunctive combination rule

For all $X \subseteq \Omega, X \neq \emptyset$, the conjunctive rule on X is given by:

$$m_{Conj}(X) = \sum_{Y_1 \cap Y_2 = X} m_1(Y_1) m_2(Y_2) \tag{2.20}$$

The conjunctive operator is denoted as $\textcircled{\cap}$ in this thesis.

Dempster's combination rule

Dempster's rule is a normalized version of conjunctive rule, given by:

$$m_D(X) = \frac{1}{1-\kappa} m_{Conj}(X) \tag{2.21}$$

where

$$\kappa = \sum_{A \cap B = \emptyset} m_1(A) m_2(B)$$

κ is generally called (global) conflict. From the concept of simple BBA and conjunctive combination rule $\textcircled{\cap}$, the definition of separable BBA is given as:

Definition 2.3.14. Separable BBA: A (normalized) BBA is separable if it can be written as the combination of simple BBAs

$$m = \textcircled{\cap}_{\emptyset \neq X \subset \Omega} X^{w(X)} \tag{2.22}$$

with $0 \leq w(X) \leq 1$ for all $X \subset \Omega, X \neq \emptyset$

Large Number of Sources (LNS) conjunctive rule

When the number of sources is large and BBAs are separable, if these sources have similar reliability, the conjunctive rule may converge to total conflict, making the rule inapplicable. LNS rule [ZMP17] solves such problems by clustering BBAs and adding proper reliability on sources.
The procedure of LNS is as follows:

1. Cluster the simple BBAs into c groups based on their focal element;
2. Combine the BBAs in the same group;
3. Reliability-based discounting;
4. Global combine the fused BBAs in different groups.

After step 1, a simple BBA from source s_j in cluster k is denoted by $X_k^{w_j}$, the rule is given by:

$$m_{LNS} = \textcircled{\cap}_{k=1,\ldots,c} (X_k)^{1-\alpha_k+\alpha_k \prod_{j=1}^{N_k} w_j} \tag{2.23}$$

where N_k denotes the number of simple BBAs in cluster k, α_k the discounting coefficient, given by:

$$\alpha_k = \frac{N_k}{\sum_{i=1}^{c} N_i} \tag{2.24}$$

When information sources are not cognitively independent, mean value combination rule is often applied, defined as follows.

Mean value combination rule

Mean value rule is often applied when sources are cognitively dependent. Given multiple sources $\mathbf{S} = \{s_1, \ldots, s_s\}$ with BBAs $m_1, \ldots, m_s$ the mean value rule is given by:

$$m(X) = \frac{1}{s} \sum_{i=1}^{s} m_i(X), \forall X \in 2^\Omega \tag{2.25}$$

Discounting

Discounting coefficients [Sha76] are applied when the reliability of sources are different. With a coefficient $\alpha_j \in [0, 1]$ on source s_j, given the BBA on s_j, noted as m_j, the discounted BBA m_j^α is defined as

$$\begin{aligned} m_j^\alpha(X) &= \alpha_j m_j(X), \forall X \subseteq \Omega \setminus \Omega \\ m_j^\alpha(\Omega) &= 1 - \alpha_j(1 - m_j(\Omega)) \end{aligned} \tag{2.26}$$

$\alpha_j = 0$ implies that the source s_j is completely unreliable and the discounted BBA represents total ignorance. It should be noted that the discounting is a pre-combination treatment method depending on the reliability of sources, rather than combination rules.

Many other combination rules exist depending on specific conditions, assumptions or properties. For example, cautious rule [Den06] is well accepted when the property of idempotence is required. Yager's rule [Yag87] assumes that the global conflict comes from the ignorance. Dubois and Prade's [DP88] rule assumes that partial conflict comes from partial ignorance. When the reliability of sources are unknown and at least one source is reliable, the disjunctive rule is often applied. Mix rules [MO07] are sometimes used for dealing with partial conflicts.

2.3.2 Decision making in TBF

In the theory of belief functions, the decision making procedure is to choose one set X in scope of the frame Ω.

In the theory of belief functions, the decision making process is to select one element $X \in 2^\Omega$. Normally, the decision is based on the evidence degree.

Credal based decision

Usually, decisions are made based on the maximum values of evidence, or credal value. The pessimist way is to decide on the singleton with maximum belief degree, ***i.e.*** singleton with maximum support from evidence, formally:

$$max_{\omega \in \Omega} Bel(\omega) \tag{2.27}$$

The optimist way is to decide on the singleton with maximum plausibility degree, ***i.e.*** singleton which is most consistent with evidence, formally:

$$max_{\omega \in \Omega} Pl(\omega) \tag{2.28}$$

A well accepted view is to manage the conflicts after the combination step [Sme07], a compromised method is given on the maximum value of pignistic probability

$$max_{\omega \in \Omega} BetP(\omega), \tag{2.29}$$

defined by:

$$BetP(X) = \sum_{Y \subseteq \Omega, Y \neq \emptyset} \frac{|X \cap Y|}{|Y|} m(Y) \tag{2.30}$$

where $|X|$ is the cardinality of set X.

Distance based decision

Decision on distance was originally proposed in [EMSY14], given a combined BBA m the decided element ω' is selected by the following rule:

$$\omega' = \underset{X \subseteq \Omega}{argmin} \{d(m, X^0)\} \tag{2.31}$$

where X^0 is a categorical BBA on $X \subseteq \Omega$, $d(m_1, m_2)$ the distance function on BBAs m_1, m_2 (such as Jousselme distance [JGEB01]).

This distance based decision rule allows to decide on all elements in 2^Ω, including partial ignorance events.

2.4 Conclusion

In this chapter, we introduced different popular models on precise and imprecise preference relations, as well as the definitions concerned. A synthetic conclusion on the definitions of preference orders is given in Tables 2.1 and 2.2, where a, b, c, d, e and f are alternatives, P the set of strict preference relations, I the set of indifference relations, $g(\cdot)$ an utility function on alternatives, and R the corresponding relation.

The adjacency matrix of total order can be transferred into a triangular matrix U.

Table 2.1 – Total & weak order

	Total Order	Weak Order
Definition	Complete Antisymmetric Transitive	Complete Transitive
Numerical	$aRb \Leftrightarrow g(a) \geq g(b)$ $g(a) = g(b) \Rightarrow a = b$	$aRb \Leftrightarrow g(a) \geq g(b)$
Properties	P transitive & complete P negatively transitive $P \cdot P \subset P$	P transitive I transitive $I \cdot P \subset P$ $P \cdot I \subset P$
Matrix Example	$\begin{array}{c\|cccccc} & a & b & c & d & e & f \\ \hline a & 1 & 1 & 1 & 1 & 1 & 1 \\ b & 0 & 1 & 1 & 1 & 1 & 1 \\ c & 0 & 0 & 1 & 1 & 1 & 1 \\ d & 0 & 0 & 0 & 1 & 1 & 1 \\ e & 0 & 0 & 0 & 0 & 1 & 1 \\ f & 0 & 0 & 0 & 0 & 0 & 1 \end{array}$	$\begin{array}{c\|cccccc} & a & b & c & d & e & f \\ \hline a & 1 & 1 & 1 & 1 & 1 & 1 \\ b & 1 & 1 & 1 & 1 & 1 & 1 \\ c & 1 & 1 & 1 & 1 & 1 & 1 \\ d & 0 & 0 & 0 & 1 & 1 & 1 \\ e & 0 & 0 & 0 & 0 & 1 & 1 \\ f & 0 & 0 & 0 & 0 & 0 & 1 \end{array}$
Permutation result	$[U]$	$\begin{bmatrix} J & J \\ 0 & U \end{bmatrix}$

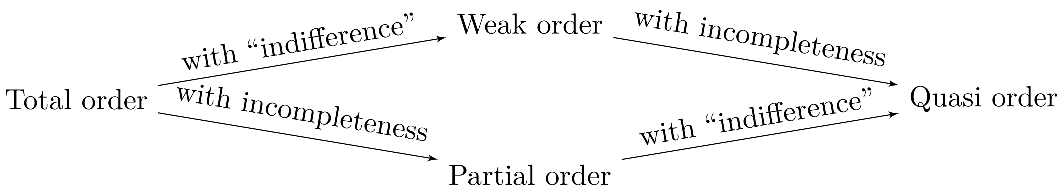


Figure 2.2 – Relations between different type of orders

For weak order, the existence of “indifference” relation makes the sub-matrix made up of ones, denoted by J, after the permutation operations.

On the conceptions of preference structures, following relations are true, illustrated in Figure 2.2. The upper relation chain is enriched with the relation of “indifference” and “incomparability” by turn while the lower one with the “incomparability” and “indifference”.

On the modeling of imprecision preferences, three theories for uncertain information are introduced, with capability of information reasoning illustrated in Table 2.3. The theory of belief functions is a mathematical tool for reasoning knowledge with imprecision and ignorance and also a powerful tool for information fusion, with different combination and decision making procedures. Several popular fusion and decision rules are introduced, with applicable circumstances concluded in Table 2.4.

Table 2.2 – Partial & Quasi order

	Partial Order	Quasi order
Definition	Reflexive Anti-symmetric Transitive	Reflexive Transitive
Numerical	$aRb \Rightarrow g(a) \geq g(b)$ $g(a) = g(b) \Rightarrow a = b$	$aRb \Rightarrow g(a) \geq g(b)$
Property	$\cap$ of a total order	P transitive I transitive $P \cdot I \subset P$ $I \cdot P \subset P$
Matrix Example	$\begin{array}{c\|ccccc} & a & b & c & d & e \\ \hline a & 1 & 0 & 1 & 1 & 1 \\ b & 0 & 1 & 1 & 1 & 1 \\ c & 0 & 0 & 1 & 1 & 1 \\ d & 0 & 0 & 0 & 1 & 1 \\ e & 0 & 0 & 0 & 0 & 1 \end{array}$	$\begin{array}{c\|ccccc} & a & b & c & d & e \\ \hline a & 1 & 1 & 0 & 1 & 1 \\ b & 1 & 1 & 1 & 1 & 1 \\ c & 0 & 0 & 1 & 1 & 1 \\ d & 0 & 0 & 0 & 1 & 1 \\ e & 0 & 0 & 0 & 1 & 1 \end{array}$

Table 2.3 – Modeling of uncertainty for different theories

Theory	Uncertainty	Imprecision	Ignorance
Probability	$\surd$	x	x
Fuzzy set	$\surd$ (intensity degree)	x	x
Possibility	$\surd$	$\surd$ (plausibility in TBF)	$\surd$
Belief function	$\surd$	$\surd$	$\surd$

Of course, many other combination rules exist, meeting different properties of combinations. In this thesis, solely the rules above are applied. To conclude, in this chapter, we introduced basic concepts on preference and theory of belief functions, which are fundamental elements for the work of this thesis. In next chapters of this part, we will introduce state-of-the-art methods on preference aggregation and learning.

Table 2.4 – Application circumstances of several combination rules

Combination rule (or process)	Circumstances
Dempster's rule	Cognitively independent and equally reliable
Mean rule	Cognitively dependent
LNS rule	Large number of simple support BBAs
Discounting rule	different reliability

Chapter 3

Similarity measures on preferences and on the theory of belief functions

Similarity measures between preferences are often applied in many preference based application such as preference aggregation and preference learning, which are introduced in Chapter 4. In this chapter, we introduce some popular methods of similarity measure applicable to preference models as well as similarity methods in the theory of belief functions.
The chapter is structured as follows: in Section 3.1, properties of distances and definitions of all types of metrics are introduced. In Section 3.2 and 3.3, popular similarity measure methods for preference relation as well as preference structures are respectively introduced and compared. Besides, in Section 3.4, similarity measures applied in TBF are also introduced, categorized into three views: distance, divergence and conflict measure.

3.1 Properties of distances

Distances on preference relations measure the similarity between distinctive preference relation types, usually applied in group consensus rules for preference aggregation. Distances on preference relation are categorized into two types: encoded preference distance and axiomatic distance. A distance function is also called a metric, possessing certain properties.

Definition 3.1.1. distance function (or metric): A distance function (or metric) on a set of objects $\mathcal{X}$ is a function

$$d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}, \tag{3.1}$$

where $\mathbb{R}_{\geq 0}$ is the set of non-negative real numbers and for all $x, y, z \in \mathcal{X}$, the following properties are satisfied:

$$\begin{array}{lr} 1.\ d(x,y) \geq 0 & \text{non-negativity or separation axiom} \\ 2.\ d(x,y) = 0 \Leftrightarrow x = y & \text{identity of indiscernibles} \\ 3.\ d(x,y) = d(y,x) & \text{symmetry} \\ 4.\ d(x,z) \leq d(x,y) + d(y,z) & \text{subadditivity or triangle inequality} \end{array} \tag{3.2}$$

Some functions measuring similarity may not possess all the four properties, introducing other definitions of ***pseudometrics, quasimetrics, metametrics*** and ***semimetrics***.

Definition 3.1.2. Pseudo-metric: Pseudo metrics do not satisfy the property of identity of indiscernibles from metric properties.

In other words, for a pseudo-metric $d_{p-metric}$,

$$\exists x, y \in \mathcal{X}, x \neq y, d_{p-metric}(x,y) = 0 \tag{3.3}$$

Definition 3.1.3. Quasi-metric: Quasi-metrics satisfy the properties of *non-negativiey, identity of indiscernibles and triangle inequality.*

Without the property of symmetry, in quasi-metric, the order of measured objects may make difference on the result.

Definition 3.1.4. Meta-metric: Meta-metrics $d_{m-metric}$ applies property of identity of a weakened discernible from metric with $d_{m-metric}(x,y) = 0 \Rightarrow x = y$ but $x = y \nRightarrow d_{m-metric}(x,y) = 0$.

Definition 3.1.5. Semi-metric: Semi-metrics drop the property of triangle inequality from metric

Semi metric is also defined as ***similarity*** or ***dissimilarity*** in many works, possessing only non-negativity, identity of indiscernible and symmetry properties. Dissimilarity is the reciprocal to similarity. In this thesis, we accept the term of similarity.

Pre-metrics are very relaxed definition, possessing solely non-negativity and idempotent. formally:

Definition 3.1.6. Pre-metric A pre-metric on $\mathcal{X}$ is a function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$, satisfying the properties of:

$$\begin{array}{l} 1.\ d(x,y) \geq 0 \\ 2.\ d(x,x) = 0 \end{array} \tag{3.4}$$

In this terminology of this thesis, distances are metrics by default and ***similarity*** possess at least the properties of semi-metrics.

The similarity measure for preferences are based on the definitions of different types of metrics above. Similarity measures for preferences fundamentally depend on the preference representation. Generally, these similarity measures are categorized into two groups: similarity between preference relation types and similarity between preference structures. As definition ***distance*** possesses most of the properties, the following introductions are mainly on different types of distances.

3.2 Distances between preference relation types

Given the preference relation types $\mathcal{R} = \{\succ, \prec, \approx, \sim\}$, distance between preference relation types measure the similarity between $R_1, R_2 \in \mathcal{R}$.

3.2.1 Distances on encoded preferences

The four preference relation types can be encoded based on adjacency matrix. Given two alternatives a_i, a_j, the preference relation types are respectively expressed by adjacency in Table 3.1.

Table 3.1 – Encoding of preference relations

Relation:	$a_i \succ a_j$			$a_i \prec a_j$			$a_i \approx a_j$			$a_i \sim a_j$		
		a_i	a_j		a_i	a_j		a_i	a_j		a_i	a_j
	a_i	0	1	a_i	0	0	a_i	0	1	a_i	0	0
	a_j	0	0	a_j	1	0	a_j	1	0	a_j	0	0
Encode:	0100			0010			0110			0000		

We denote encoding function as $enc(\cdot)$, and the four bits of encoded preference as $enc_{1-4}()$, the encoded preferences are in form of vectors. Thus, Minkowski distance [JD88] can be applied. Given two preference relation types $R_1, R_2 \in \{\succ, \prec, \approx, \sim\}$, the Minkowski distance $d_{minkowski}$ is defined as:

$$d_{minkowski}(R_1, R_2) = \left(\sum_{i=1}^{4} |enc_i(R_1) - enc_i(R_2)|^p \right)^{\frac{1}{p}} \tag{3.5}$$

where p denotes the norm L^p (Lebesgue space).
For encoded data, Manhattan distance is usually applied, ***i.e.*** $p = 1$ in Minkowski distance. Thus, the Manhattan distance on preference relation types is given in Table 3.2. Manhattan distance is a measure function on encoded data, however, it is not able to keep some intrinsic properties of preferences. For instance, there is no reason supporting that $d(\succ, \prec) = d(\approx, \sim)$. To keep such properties, axiomatic distances were proposed.

Table 3.2 – Hamming distance (Manhattan distance) between two preference relation types

	$\succ$	$\prec$	$\approx$	$\sim$
$\succ$	0	2	1	1
$\prec$	2	0	1	1
$\approx$	1	1	0	2
$\sim$	1	1	2	0

3.2.2 Axiomatic distances for preferences

Axiomatic distances are based on appropriate axioms on preference relations, with also properties of metrics respected. Such distances were initiated by Kemeny and Snell in 1963 [KS63] for consensus aggregation rule on weak orders (a.k.a. KS model), then generalized by Bogart [Bog73, Bog75] to deal with partial orders, which excludes the "indifference" relation. Afterwards, Cook ***et. al.*** [CKS86] developed a distance on bi-matrix representation of quasi orders (a.k.a. CS model). Extended from KS model, J. M. Blin [Bli76] proposed an aggregation method by distance maximization procedure. Khelifa and Martel [KM01] proposed an axiomatic distance by taking the center value in a restrained space limited by axioms for consensus group decision making by respecting RS model [RS93], namely KM model. These various models do not accept identical axioms, causing some differences in numeric values of similarities.

CS distance follows the axioms demonstrated in Figure 3.1. In this axiom, the inverse strict preference relations are considered as the most distant. "indifference" is between the "incomparability" and strict preference. Thus the distances respect the following relation:

$$d(\sim,\succ) = d(\sim,\prec) = d(\sim,\approx) + d(\approx,\succ) \tag{3.6}$$

With the minimum positive distance valued as 1, the CS distance is given in Table 3.3.

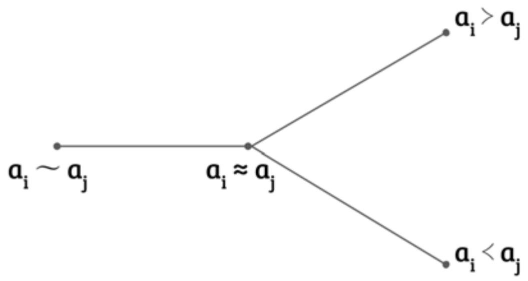


Figure 3.1 – Distance relation accepted in CS distance

In RS model, the distances are values following the Figure 3.2. The distance between "incomparability" are abstractly defined by x and y without giving a precise value.

Therefore, the distances between preference relations are given in Table 3.4.

Table 3.3 – Distance between different preference relation types in CS distance

	$\succ$	$\prec$	$\approx$	?
$\succ$	0	2	1	2
$\prec$	2	0	1	1
$\approx$	1	1	0	2
?	2	2	1	0

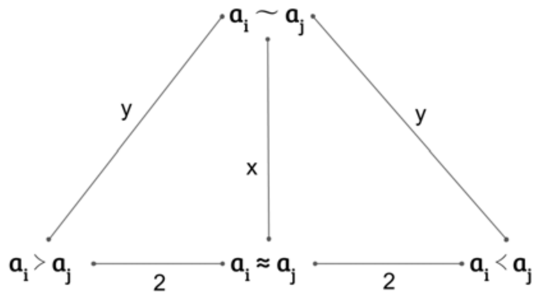


Figure 3.2 – Distance relation accepted in RS distance.

According to the axioms in [RS93], a synthetic relation between x, y is limited in Equation (3.7).

$$0 < max\{2, x\} \leq y \leq min\{4, 2 + x\} \tag{3.7}$$

In KM distance, the authors regard the "incomparability" as an absence of knowledge and applied the insufficiency principle of Laplace. Thus, the distance from "incomparability" to other relations are equivalent, formally:

$$d(\sim, \succ) = d(\sim, \prec) = d(\sim, \approx) \tag{3.8}$$

With minimum distance valued as 1, the numeric distance between different preference types are given in Table 3.5. The detail of the calculation is given in [KM01].

More discussion on axiomatic preference distances are given in Section 3.5, with a synthetic comparison between different distance models summarized.

3.3 Distances in preference structures

Preference structures as total orders (or linear order) can be regarded as ordinal variables. Thus, rank correlation coefficients and distances can be applied. Some most popular rank correlation distances are:

- Spearman's footrule distance [DG77]
- Kendall's τ [Ken48]
- Fagin's distance [FKM$^+$04] for weak orders

Table 3.4 – Distance between different preference relation types in RS distance.

	$\succ$	$\prec$	$\approx$	$\sim$
$\succ$	0	4	2	y
$\prec$	4	0	2	y
$\approx$	2	2	0	x
$\sim$	y	y	x	0

Table 3.5 – Distance between different preference relation types in KS model

	$\succ$	$\prec$	$\approx$	?
$\succ$	0	$\frac{5}{3}$	1	$\frac{4}{3}$
$\prec$	$\frac{5}{3}$	0	1	$\frac{4}{3}$
$\approx$	1	1	0	$\frac{4}{3}$
?	$\frac{4}{3}$	$\frac{4}{3}$	$\frac{4}{3}$	0

It should be noted that correlation coefficients are not distances since the coefficient function values are in range of $\lfloor -1, +1 \rceil$ rather than $\mathbb{R}_{\geq 0}$.

For two total preference orders σ_1, σ_2 in same alternative space $\mathcal{A} = \{a_1, \ldots, a_n\}$

$$\sigma_1 : a'_1 \succ a'_2 \succ \ldots \succ a'_n$$

and

$$\sigma_2 : a''_1 \succ a''_2 \succ \ldots \succ a''_n$$

denote the ranking function as $rank(a)$, indicating the ranking position of the alternative a in the corresponding order. The different correlation coefficients and distances are defined as follows:

Spearman's footrule distance

Spearman's footrule distance $d_{spearman}$ measure the similarity between two ranks by number of displacement (to make them equal). Defined by:

$$d_{spearman}(\sigma_1, \sigma_2) = \sum_{i=1}^{n} (rank_{\sigma_1}(a_i) - rank_{\sigma_2}(a_i)) \tag{3.9}$$

Spearman's footrule distance is also a Manhattan distance for rank variables.
Spearman's correlation coefficient ρ is a coefficient varying from -1 to +1, where +1 implies the identical case and -1 the inverse case, defined as Spearman's ρ:

$$\rho(\sigma_1, \sigma_2) = 1 - \frac{6 \sum_{i=1}^{n} (rank(a_i) - rank(a'_i))^2}{n(n^2 - 1)} \tag{3.10}$$

Kendall's τ and Fagin's distance

In Kendall's τ, alternatives are compared in a pairwise way. For any pair of alternatives a_1, a_2, between two orders σ_1, σ_2 denote $a_1 R_\sigma a_2$ for relations between alternatives in order σ, with $R \in \{\succ, \prec\}$. We give the definition of ***concordant*** and ***discordant***, with a discordant function on alternatives a_i, a_j in orders σ_1 and σ_2 denoted as $\bar{K}_{i,j}(\sigma_1, \sigma_2)$:

- If $a_1 R_{\sigma_1} a_2 = a_1 R_{\sigma_2} a_2$, a_1 and a_2 are concordant in σ_1 and σ_2, thus, $\bar{K}_{i,j}(\sigma_1, \sigma_2) = 0$;
- If $a_1 R_{\sigma_1} a_2 \neq a_1 R_{\sigma_2} a_2$, a_1, a_1 and a_2 are discordant in σ_1 and σ_2, thus, $\bar{K}_{i,j}(\sigma_1, \sigma_2) = 1$;
- If $a_1 \approx_{\sigma_1} a_2$ and $a_1 \approx_{\sigma_2} a_2$, a_1 and a_2 are neither concordant nor discordant.

The Kendall's distance $d_{Kendall}$ is simply the number of discordant pairs, more formally:

$$d_{Kendall}(\sigma_1, \sigma_2) = \sum_{i<j} \bar{K}_{i,j}(\sigma_1, \sigma_2), \tag{3.11}$$

and Kendall's correlation coefficient τ is defined as:

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{n(n-1)/2} \tag{3.12}$$

Fagin's distance [FKM+04] is an extension version of Kendall's τ, adapting to weak orders, i.e. the "indifference" relation is considered. Extended from the discordant function above, Fagin's distance considers the following case:

- if $a_1 \approx_{\sigma_1} a_2$ and $a_1 R_{\sigma_2} a_2$, the discordant is values by the penalty p, i.e. $\bar{K}_{i,j}(\sigma_1, \sigma_2) = p$.

According to [FKM+04], the choice of p must respect:

$$\frac{1}{2} < p < 1 \tag{3.13}$$

The calculation of Fagin's distance is similar to Kendall's τ with the new case considered:

$$d_{Fagin}(\sigma_1, \sigma_2) = \sum_{i<j} \bar{K}_{i,j}(\sigma_1, \sigma_2), \tag{3.14}$$

Diaconis and Graham proved that the Spearman's footrule and Kendall's τ are robust [DG77]: For any ordinal rankings σ_1, σ_2 on the same alternative space A

$$d_{Kendall}(\sigma_1, \sigma_2) < d_{Spearman}(\sigma_1, \sigma_2) < 2 \times d_{Kendall}(\sigma_1, \sigma_2) \tag{3.15}$$

3.4 Similarity measures in the theory of belief functions

In the theory of belief functions, different similarity measures are also applied on BBAs. From geometrical and statistical views, these measures are categorized into three parts: distance, divergence and conflicts[1].

3.4.1 Geometry view: distance

As BBAs are defined in space of 2^Ω, In addition to all the properties for metrics in Equation (3.2), properties considering structure of discernment are specific to BBAs. [JM12] consider following properties:

- Strong structural property:
  A distance measure d between two BBAs m_1 and m_2 is strongly structural if its definition accounts for the interaction between the focal elements of m_1 and m_2.
- Weak structural property:
  A distance measure d between two BBAs m_1 and m_2 is weakly structural if its definition accounts for the cardinality between the focal elements of m_1 and m_2.
- Structural similarity:
  A distance measure d between two BBAs m_1 and m_2 is structural similarity if its definition accounts for the interaction between the set $\mathcal{F}_1$ and $\mathcal{F}_2$ of focal elements of m_1 and m_2.

A popular distance that is strongly structural is Jousselme distance [JGEB01], defined as:

$$d_{Jousselme}(m_1, m_2) = \sqrt{(\boldsymbol{m_1} - \boldsymbol{m_2})^T \boldsymbol{Jacc}(\boldsymbol{m_1} - \boldsymbol{m_2})} \tag{3.16}$$

where $\boldsymbol{Jacc}$ is the matrix whose elements are Jaccard indices:

$$Jacc(X_1, X_2) = \frac{|X_1 \cap X_2|}{|X_1 \cup X_2|}, \text{for } X_1, X_2 \in 2^\Omega \setminus \emptyset \tag{3.17}$$

These properties are "natural" to the distances on BBAs, however, the importance or advantages have not been theoretically proved. Some other properties are proposed in the contribution part of the thesis

3.4.2 Statistic view: divergence

As BBAs express also the imprecision information, they can be measured from a view of entropy. Divergence is used for similarity measure in entropy, such as Kullback-Leibler divergence [KL51] for probability distribution. Perry and Stephanou proposed

[1] Strictly speaking, "conflict" is not a similarity measure (semi-metric). Since it measures the consistence of different BBA, we put it in this section.

a divergence d_{PS} for BBAs by applying Dempster's combination rule rather than Bayes' rule:

$$d_{PS}(m_1, m_2) = |\mathcal{F}_1 \cup \mathcal{F}_2| \left(1 - \frac{|\mathcal{F}_1 \cap \mathcal{F}_2|}{|\mathcal{F}_1 \cup \mathcal{F}_2|}\right) + (\boldsymbol{m_1} \textcircled{\cap} \boldsymbol{m_2} - \boldsymbol{m_1})^T (\boldsymbol{m_1} \textcircled{\cap} \boldsymbol{m_2} - \boldsymbol{m_2}) \tag{3.18}$$

where $\textcircled{\cap}$ is the Dempster's combination operator. d_{PS} has two components respectively measuring the structural similarity and information change.

For futher reading, a synthetic study on distances and divergence in TBF is done in [JM12].

3.4.3 Conflicts measuring

As BBAs are able to express ignorance, the term conflict is used to describe the coherence between evidential knowledge, especially those with partial ignorance. The most common one is ***global-conflict*** $Conf(\cdot)$, defined on the Dempster's combination rule:

$$Conf(m_1, m_2) = (m_1 \textcircled{\cap} m_2)(\emptyset) \tag{3.19}$$

where $\textcircled{\cap}$ denotes the Dempster's combination operator.
Indeed, global-conflict measures the contradiction between BBAs. In [DB13], the authors proposed and summarized several properties needed for measuring conflicts between BBAs.Conflicts are also based on other measures. In [MJO08], authors proposed a conflicting measure method developed from Jousselme distance. Another conflict measure based on the inclusion is proposed in [Mar12]. A global introduction on conflict management is given in [Mar19]. The conflicts are usually applied in decision making procedures. In fact, the application on conflicts in TBF is limited than distances because they may not be metric.

3.5 Conclusion

In this chapter, we illustrated the basic notions on metrics and reviews different distances for preference relations and structures. Some conflicting axioms on axiomatic distances are compared and discussed. For distances on preference relations, all distances respect that the distance between "indifference" and "strict preference" is less than that between two inverse strict preferences. The most controversial issue focuses on the similarity of "incomparability" with other preference relation types.

While for Minkowski distance (including Manhattan distance), the relations are defined by encoded formats with "incomparability" defined as an exclusive concepts to other relation types. The distances between four preference relations are illustrated in Figure 3.3.

For example, CS model, KM (JMK) model and RS models have different axioms concerning the relation of "incomparability". In fact, in RS model [RS93], the "incomparability" is defined as:

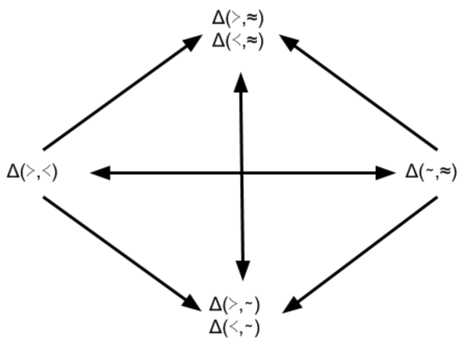
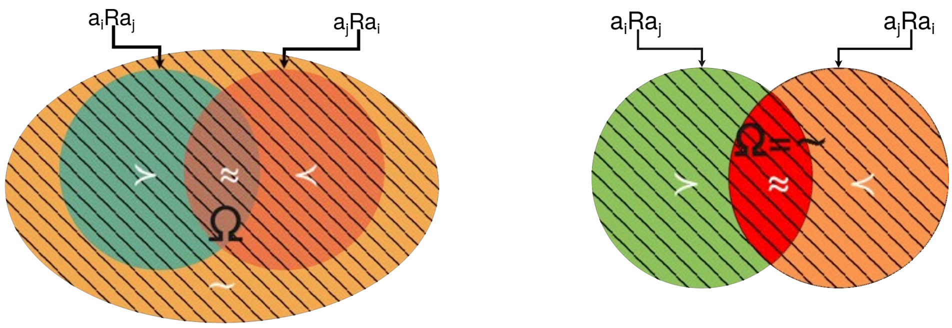


Figure 3.3 – Preference distance applied in Minkowski family.
$a \longrightarrow b$ implies $a > b$
$a \longleftrightarrow b$ implies $a = b$

> Incomparability relation is the affirmation of the incapacity to establish the relation type: there is no indifference, no weak preference and no strict preference between the two alternatives.

Such definition indicates the illustration in Figure 3.4a. In CS model and KM (JMK) models, the incomparability is defined as "the two alternatives are not compared", corresponding to the definition in Figure 3.4b.



(a) "Incompleteness" caused by ignorance

(b) "Incompleteness" caused by ignorance without "incomparability"

Figure 3.4 – Venn diagrams of the preference relations

In order to distinguish this disagreement on the definition of incompleteness, we express the relations by $\{\succ, \prec, \approx, \sim, ?\}$, where $\langle \succ \rangle$ denotes the "incomparability" relation and $\langle ? \rangle$ the missing information. The accepted axioms are listed as follows:

1. Axioms 1: properties of metric are satisfied:
   (a) Axiom 1a (Non-negativity): $\forall R_1, R_2 \in \{\succ, \prec, \approx, \sim\}, \Delta(R_1, R_2) \geq 0$, with equality if and only if $R_1 = R_2$
   (b) Axiom 1b (identity of indiscernibles): $\Delta(R_1, R_2) = 0 \Longleftrightarrow R_1 = R_2$

(c) Axiom 1c (Symmetry): $\forall R_1, R_2 \in \{\succ, \prec, \approx, \sim\}, \Delta(R_1, R_2) = \Delta(R_2, R_1)$

(d) Axiom 1d (Triangle inequality): $\forall R_1, R_2, R_3 \in \{\succ, \prec, \approx, \sim\}, \Delta(R_1, R_3) + \Delta(R_3, R_2) \leq \Delta(R_1, R_2)$

2. Axiom 2: $\Delta(\succ, \approx) = \Delta(\prec, \approx)$ and $\Delta(\succ, \sim) = \Delta(\prec, \sim)$
(BIS: $\Delta(\succ, ?) = \Delta(\prec, ?)$). This axiom says $\succ$ and $\prec$ are opposite relations

3. Axiom 3: $\Delta(\succ, \approx) + \Delta(\approx, \prec) = \Delta(\succ, \prec)$.

4. Axiom 4: $\Delta(\succ, \sim) \leq \Delta(\approx, \sim)$
(BIS 1: $\Delta(\succ, ?) \leq \Delta(\approx, ?)$
BIS 2: $\Delta(\succ, ?) = \Delta(\approx, ?)$)

5. Axiom 5: $\Delta(\approx, \sim) \leq \Delta(\approx, \succ)$
(BIS: $\Delta(\approx, ?) \leq \Delta(\approx, ?)$)

6. Axiom 6: $\Delta(\succ, \prec) = max(\{\Delta(R_1, R_2) : R_1, R_2 \in \{\succ, \prec, \approx, \sim\}\})$,
meaning that strict preference and inverse preference relations are mostly distinguished.

For the axiomatic models mentioned above, the distance relations are illustrated in Figure 3.5.

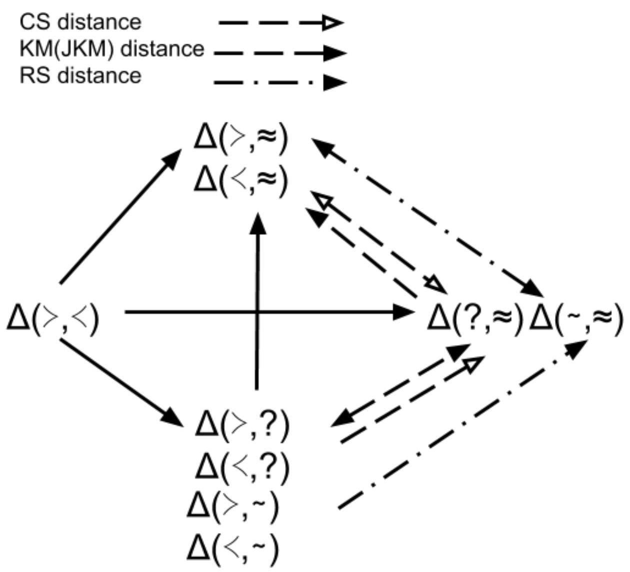


Figure 3.5 – Preference distance applied in three different models, with:
$a \longrightarrow b$ implies $a > b$
$a \longleftrightarrow b$ implies $a = b$

The comparison demonstrated in Table 3.6 sums up the controversial part between four types of distances, especially concerning the "incomparability" relation.

Table 3.6 – Controversial similarity related to "incomparability" in different relation models.

Distance	Distance to incomparability
Mincowski distance	$d(\succ,\sim) = d(\succ,\approx)$ $d(\succ,\prec) = d(\approx,\sim)$
CS distance	$d(\succ,?) > d(?,\approx)$ $d(\succ,\approx) = d(?,\succ)$ $d(\succ,\prec) > d(?,\approx)$
JKM distance	$d(\succ,?) = d(?,\approx)$ $d(\succ,\approx) < d(?,\succ)$ $d(\succ,\prec) > d(?,\approx)$
RS distance	$d(\succ,\sim) > d(\sim,\approx)$ $d(\succ,\approx) = d(\sim,\succ)$ $d(\succ,\prec) > d(\sim,\approx)$

To recapitulate, there is no standard distance between preference relation types accepted by all researchers, and the definition of "incompleteness" remains a disputable issue. The choice of distance on preference relations is a flexible process. More arguments on this issue is contributed in Section 5

We also introduced similarity measure in TBF. As TBF is a tool for uncertainty modeling, more criteria are applicable in addition to distance. Three genres of methods can be categorized, defined from different aspects of view, resumed in Table 3.7.

Table 3.7 – Points of view on similarity measure between BBAs.

Similarity	Aspect
Distance	Geometry: Each BBA is represented by a vector. Thus distance for vectors are applicable
Divergence	Statistic: A BBA represent the imprecision of the information source. Thus entropy is applicable.
Conflict	Consistency: A BBA represent the ignorance of the information source. Thus conflict is applicable considering the ignorance factor.

In deed, similarity measures are useful in preference management including preference aggregation and preference learning, introduced in more details in next chapters.

Chapter 4

Preference management

In this chapter, we review state of the art methods on preference management problems. In the scope of this thesis, we mainly studied two important sub-domains of preference management problems, preference aggregation and preference learning. Generally speaking, preference aggregation is the procedure of merging multi-agents' preferences into one. Preference learning is a combined domain of preference modeling and machine learning, either by applying machine learning methods on preference data, or by applying preference concepts in machine learning algorithms design. The chapter is divided into two parts: in Sections 4.1, 4.2, 4.3, and 4.3.2, aggregation problems and methods are introduced. In Sections 4.4 a global view on preference learning is introduced, followed by preference clustering section (Section 4.5).

4.1 Preference aggregation functions

Preference aggregation problem is also called "social choice problem" in the study of sociology and economy. Briefly speaking, preference aggregation serves to merge multiple preferences on same group of alternatives, and finally result into one alternative. The problem setting is defined as follows.
Social choice setting: A preference aggregation problem setting (or ***social choice setting***) is defined by a triple $(\mathcal{A}, \mathcal{AGT}, \mathcal{OD})$ where:

1. $\mathcal{A} := \{a_1, \ldots, a_K\}$ of size K is a finite set of outcomes, also called candidates, or alternatives.

2. $\mathcal{AGT} := \{agt_1, ..., agt_N\}$ of size N is a finite set of agents

3. $\mathcal{OD} := \{\sigma_1, \ldots, \sigma_M\}$ of size M is a set of preferences ordering over $\mathcal{A}$

Preference aggregation methods is an application in social choice theory. As this thesis is not mainly on the topic of sociology, the utilization of the term "social choice" may cause ambiguity. Hence, we insist on the term "preference aggregation".

Definition 4.1.1. Preference aggregation function (social choice function): A preference aggregation function over $\mathcal{AGT}$ and $\mathcal{A}$ is a function $C : \mathcal{OD}^n \rightarrow \mathcal{A}$.

When there are only two alternatives, the most widely used voting system, and arguably the most natural, is majority rule. Under majority rule, we take the alternative that is preferred by a majority of the voters and rank it first, placing the other alternative second. For this discussion we will assume that the number of voters is odd, so that we do not need to worry about the possibility of majority rule producing ties. Since majority rule is so natural in the case of two alternatives, it is natural to try designing a voting system based on majority rule when there are more than two alternatives. This, however, turns out to be remarkably tricky. The most direct approach is to first create group preferences, by applying majority rule to each pair of alternatives, and then trying to turn these group preferences into a group ranking. That is, we create a group preference relation $\succ$ out of all the individual preferences $\succ_k$ as follows. For each pair of alternatives a_i and a_j, we count the number of individuals for whom $a_i \succ_k a_j$ and the number of individuals for whom $a_j \succ_k a_i$. If the first number is larger than the second, then we say that the group preference $\succ$ satisfies $a_i \succ a_j$, since a majority of the voters prefer a_i to a_j when these two alternatives are considered in isolation. Similarly, we say $a_j \succ a_i$ in the group preference if $a_j \succ_k a_i$ for a majority of the individuals k. Since the number of voters is odd, we cannot have equal numbers favoring a_i and favoring a_j. Hence, for every distinct pair of alternatives we will have exactly one of $a_i \succ a_j$ or $a_j \succ a_i$. That is, the group preference relation is complete.

4.2 Voting paradoxes and impossibility theorems

In this section, we list some of the most famous paradoxes and impossibility theorems. It should be pointed that in the work of this thesis, only Condorcet's paradox is addressed. Other issues are for reading references.

Condorcet's paradox

Various paradoxes and impossibility theorems may arise from a preference fusion process with majority rule, and one of the most well-known is Condorcet's paradox, also known as voting paradox [dCmdC85]. It is a situation in which collective preferences can be cyclic, (***i.e.*** not transitive) even the preferences of individual voters are transitive. The avoidance of Condorcet's Paradox is also referred to as guarantee of "consistence" or "transitivity" in some work.

Other paradoxes are also studied in decision making theory, such as Borda's paradox, Ostrogorski's paradox [RD76], the referendum paradox, ***etc***. As these paradoxes are not in the scope of our work, detailed definitions are not given here. Some others problems exist concerning more specific issue corresponding to the circumstances. For example, in voting systems for political regime, it is better to guarantee that the elected candidate get the majority votes for the sake of stability, even though plural (more than two) candidates participate in the campaign.

4.3 Aggregation on various preference models

In this section, we introduce several aggregation rules on both crisp preference and fuzzy preference. For crisp preference, some popular traditional strategies as well as distance-based strategy are introduced. For fuzzy preference, we briefly introduce the Operator Weight Average (OWA) method as well as its extensions.

4.3.1 Aggregation on crisp preferences

Aggregation functions on crisp preferences usually based on voting rules. We introduce some of the most popular voting rules, depending on different principles.

Condorcet methods

A Condorcet method [dCmdC85] is an aggregation rule that the aggregated alternative is preferred by the majority to all other alternatives. Such alternative is also called the ***Condorcet winner***. As aforementioned, due to Condorcet's paradox, such alternative do not exist when the preference relation among multiple alternative create a circle.

Borda count

In Borda count [Eme13], agents rank alternatives in order of preference, with a number of points given to each alternative corresponding to its ranking position. A higher ranked alternative obtains a bigger number of points. Once all agents have been counted the alternative with the most points is the winner.
The Borda count is often described as a consensus-based voting system rather than a majoritarian one [Lip13], and it's able to determine multiple winners.

Plurality runoff

In plurality voting, each agent is allowed to vote for only one alternative, and the one who polls the most among its counterparts is elected. This method is easy to practice for plural alternatives. However, plurality voting is not a majoritarian voting, making it unable to reflect the consensus of the entire agents. Therefore, run-off mechanism is often introduced. In this mechanism, the first two winners of plurality rule go to the second round, and agents apply a majority voting on these two alternatives.
Plurality runoff is usually applied in political elections, such as president election in France, Parliament delegates election in Iran and mayors election in Italy.
Other voting rules also exist adapting to various using circumstances. As such work is not our focus, we recommend to refer to [Lip13] for further reading.

Pairwise distance based preference aggregation

Many preference aggregation rules are designed on the minimization of distances between alternative pairs also called distance based consensus rules. Generally, it can be concluded by Algorithm 1.

Given alternative set $\mathcal{A} = \{a_1, \ldots, a_k\}$, agents in agent set $\mathcal{AGT} = \{agt_1, \ldots, agt_n\}$ express their preferences on pairwise alternatives in A with possible preference relations in $\{\succ, \prec, \approx, \sim\}$. We denote a preference relation function between two alternatives a_i and a_j from agent agt_s as $\mathbf{Pref}^s(a_i, a_j) \to \{\succ, \prec, \approx, \sim\}$, a distance based preference aggregation procedure is defined in Algorithm 1.

Algorithm 1 Pairwise distance-based preference aggregation algorithm

Input: $\mathbf{Pref}^s(a_i, a_j)$:
  Preferences from agents $\mathcal{AGT}$ on alternatives $a_i, a_j \in \mathcal{A}$
Output: $\mathbf{Pref}^*(a_i, a_j)$:
  Aggregated preference on alternatives $a_i, a_j \in \mathcal{A}$
  Decision on every alternative pair

```
for a_i, a_j ∈ A, i < j do
    for R ∈ {≻, ≺, ≈, ∼} do
        Compute Φ_{i,j}(R) = Σ_{s=1}^{s=n} d_Δ(R, Pref(a_i, a_j))
    end for
    Decide Pref*(a_i, a_j) = {R : argmin_{R∈{≻,≺,≈,∼}} Φ_{i,j}(R)}
end for
```

$$\Phi_{i,j}(R) = \sum_{s=1}^{s=n} d_{\Delta}(R, \mathbf{Pref}(a_i, a_j))$$

$$\mathbf{Pref}^*(a_i, a_j) = \{R : \underset{R \in \{\succ, \prec, \approx, \sim\}}{argmin} \Phi_{i,j}(R)\}$$

In fact, many traditional voting rules are intrinsically the minimization of certain distances. P. Viapini [Via15] has studied this correspondence, giving the following conclusions between similarity measure and aggregation methods:

Table 4.1 – Correspondence between some distances and aggregation methods

Aggregation method	Similarity measure	Properties
Plurality	d_{PL}	premetric
Top-k	d_{TK}	premetric
Veto	d_V	premetric
Borda	Spearman	metric
Scoring rule	Positional Spearman	metric

d_{PL}, d_{TK} and d_V respectively denote similarity measures corresponding to Plurality, top-k, and veto rule.

4.3.2 Aggregation on uncertain preferences

In Section 2.2, fuzzy preference model was introduced confronting preferences with uncertainty. Many aggregation on fuzzy preferences have been proposed. In the following part, we introduce several main-stream method.
Given the fuzzy preferences from all agents, a collective fuzzy preference relation is obtained by aggregating all completed individual fuzzy preference relations. An intuitive way is to take the average values. We denote the fuzzy preference between a_i and a_j from agent agt_k as r_{ij}^k and the weight of agent agt_k as w_k, for K agents, the aggregated fuzzy preference is calculated by:

$$r_{ij} = \frac{1}{K} \sum_{k=1}^{K} w_k r_{ij}^k \tag{4.1}$$

This method is named Ordered Weighted Averaging (OWA) operator, originally proposed by [Yag88]. Almost all aggregation methods on fuzzy preferences are based on OWA. Extended from OWA, induced OWA (IOWA) operator [Yag03, YF98, YF99] has received increasing attention. In this method, the meta-data on alternatives are used to induce the ordering in the first place, for purpose to estimate values of some unknown information. To avoid the Condorcet's paradox, ***i.e.*** to keep the property of transitivity of the aggregated preference, in [HVCHA07], authors proposed additive consistency IOWA (AC-IOWA). [CHVAH08] studied the cardinal consistency and proved that it is a characterization of multiplicative transitivity. [MC11] proposed Induced Order Weighted Averaging Distance (IOWAD) operator to aggregate fuzzy preferences from a view point of distance, similar to Algorithm 1.

For futher reading, we list some applications of fuzzy preferences. Fuzzy preferences are often applied in linguistic information. [CAHV09] proposed a consensus model for unbalanced fuzzy linguistic information. [MGL13] proposed 2-TILGOWA operator to estimate linguistic values and applied for decision making. [Xu07] applied OWA on intuitionistice fuzzy preference relations (IFPR) and [pCqX19] developed aggregation method with the consideration on consistency for IFPR.

Although fuzzy preference model is powerful in expressing uncertainty[1] in preferences, it still has some shortcomings. A very important one is that it merely represent the degree (or probability) of the preference relation and could not express the imprecision problems as well as the conflicts in decision making procedures are hardly kept. A solution on these shortcomings is proposed in Chapter 5 in the contribution part of this thesis.

In addition to the aggregation, another management application on preference is preference learning.

[1]In many works, the fuzzy preference is interpreted as an expression of intensity.

4.4 Preference learning

Preference learning is a sub-field of machine learning, particularly oriented to applications and theories concerning preference knowledge. In preference management, learning on preference data is also an important aspect, often applied in preference elicitation applications. In this section, we firstly give a global introduction on preference learning with corresponding notions. Afterwards, we focus on one specific aspect of preference learning, clustering on preferences, which is one of the major topics in this thesis.

The procedure of clustering on preferences is important in both preference elicitation and community detection. Many preference elicitation methods follow the strategies that predict the unobserved preferences of an agent based on the observed preferences of neighbor agents, with more details in following contents. In social networks and sociology, preferences can be applied as a part of profile information for community detection. In this section, we firstly give a global introduction on the research domain of preference learning and its position in the study of artificial intelligence. Afterwards, we concentrate on clustering methods on preferences in Section 4.5, which is in the scope of our thesis work.

In addition to preference modeling and reasoning methods mentioned in previous chapters, preference learning is often aimed at preference elicitation (or predicting, or mining) problems. “ Roughly speaking, preference learning is about inducing predictive preference models from empirical data ” [FH11]. Methods for learning preferences in an automatic way are among the recent research topics in disciplines such as machine learning, knowledge discovery, and recommendation systems. Preference learning is strongly related to various scientific domains, a brief illustration is given in Figure 4.1, originally proposed in [FHR+14].

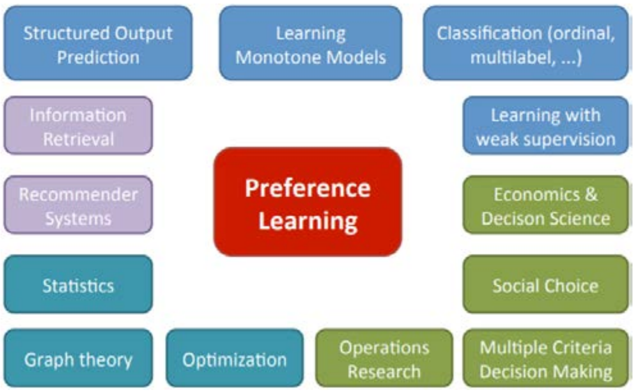


Figure 4.1 – Preference learning and related research areas within machine learning (blue), information retrieval (purple), applied mathematics (turquoise), and the decision sciences (green), from [FHR+14]

.

Preference learning is a relatively large concept. Below, we introduce the specific scenarios and techniques in preference learning. For better understanding, a topic map may be helpful in Figure 4.3 in the conclusion part of this chapter.

4.4.1 Two scenarios

According to [FH10], preference learning mainly consists of two scenarios: learning from label preferences and learning from object (or alternatives) preferences . The former one focuses on applying preference management techniques in supervised learning, and the latter one on applying machine learning techniques on preference knowledge. Formal concepts are given below.

Learning from label preferences

This scenario is also called ***label ranking*** in [FH11]. Learning from label preferences concerns the problems in supervised learning methods.
Given:

- a set of training instances $\{x_k|k = 1, \ldots, n\} \subseteq \mathcal{X}$,
- a set of labels $\mathcal{L} = \{\lambda_i|i = 1, \ldots, m\}$,
- and a set of associated pairwise preferences on labels $\lambda_i \succ_{x_k} \lambda_j$ of each training instance x_k,

learning from label preferences aims to find a mapping function of labels ranking on each alternative $\mathcal{X} \to \mathcal{M}_{XL}$, where $\mathcal{M}_{XL}$ denotes the assignment of a ranking permutation $\succ_x$ on $\mathcal{L}$ to every $x \in \mathcal{X}$. In other words, learning from label preferences is to predict the label ranking in $\mathcal{L}$ for any instance $x \in \mathcal{X}$, where $\lambda_i \succ_x \lambda_j$ implies that instance x prefers label λ_i to λ_j. This scenario is often encountered in ensemble learning, where final labels of objects are obtained by the aggregation of multiple classifiers, such as Bagging [Bre96] and Adaboost [FS97].

In deed, learning of label preferences has been a trend of extending machine learning methods to complex and structured output spaces [FH03, THJA04], notably in multi-label classification, where multiple labels may be assigned to each instance..

Learning from object preferences

This scenario is also called ***object ranking*** is other works as [FH11]. In the scenario of learning from object preferences, the machine learning algorithms are applied on objects with preference relation known to predict preference relation on new objects. This scenario is formally defined as follows.
Given

- a (potentially infinite) set $\mathcal{A}$ of objects (alternatives),

- and a finite set of pairwise preferences $a_i R a_j, (a_i, a_j) \in \mathcal{A} \times \mathcal{A}, R \in \{\succ, \prec, \approx, \sim\}$,

Methods of learning from object preferences aim to find a ranking function $\mathbf{O}(\cdot)$ that returns a permutation ranking on object $\mathcal{A}', \mathcal{A}' \subset \mathcal{A}$.

Briefly speaking, learning from object preference is the elicitation of implicit preferences based on the observed preference knowledge. This scenario is the most fundamental task in the domain of preference learning and has been widely applied, especially in recommendation systems.

In [FH11], the authors consider another scenario, ***instance ranking***, in addition to the two above. Roughly speaking, instance ranking aims to rank objects (instances) based on their corresponding label (or class) information. As instance ranking also aims to find ranking function on objects, we hereby consider it as a sub-scenario of object ranking. More details are available in [FH11].

To recapitulate, learning from object preferences is to predict preference relations or rankings by using machine learning methods while learning from label preferences is to predict preferences among labels in machine learning problems. The two scenarios are not isolated. In the scenario of learning from label preferences, labels are sometimes regarded as objects thus technique of object ranking can be applied.

In this thesis, the term “preference learning” refers to the second scenario–learning from object preferences. In the following content, major techniques for learning on object preferences are introduced.

4.4.2 Major techniques for preference learning

Preference learning techniques mainly consist of two categories: learning utility functions and learning preference relations, respectively introduced as follows.

Learning utility functions

An utility function is a mapping function $f : \mathcal{A} \to \mathbb{R}$ that assigns an utility degree $f(a)$ to each object $a \in \mathcal{A}$, which induces a linear order on $\mathcal{A}$. A ranking relation $a_i \succ a_j$ is derived by values of the utility function $f(a_i) > f(a_j)$.
A common example of utility function is the scoring (or noting) systems, such as Netflix Prize [BL$^+$07] for film scores predicting, and Sushi scores in Sushi dataset [Kam03a]. Such techniques have attracted many researchers. A most basic idea is to portray each agent by the scores on each alternative and learning directly in the space of scores. Several chanllenges exist in such methods, such as incapability to deal with missing values and low efficiency when data dimension is large. Targeted on these issues, many solutions have been proposed, such as Matrix Factorization (MF) [KP13, ZWFM06], collaborative filtering (CF) [Kam03a, CMF08, HKBR99, Paz99], Deep Learning [IJW$^+$19, ZYST19], and hybrid methods [WWY15]. More introduction on these solutions is given in Section 4.5.

Learning preference relations

However, preferences do not always come with utiliy functions. In some cases, where preferences are expressed by orders, or by pairwise relations, learning preference relations becomes necessary. Besides, [CSS98, DJBW03] have pointed out that obtaining information of preference relations may be easier and more natural than obtaining the labels needed for a classification or regression approach, and such type of information is more accurate, since " people tend to rate their preferences in a relative way, comparing objects with the other patners in the same batch", namely batch effect [DDCLB08].

Learning preference relations focuses on comparing pairs of objects (alternatives) in terms of binary preference relations. This kind of approach has been pursued in [CSS98] and applied in various domains from recommendation on internet [Joa02] to agriculture production [BBD+04]. Some important survey works are introduced in [KKA10, FH03].

Learning from object preferences has been widely applied in research works on Recommendation Systems (RS)[2], where clustering techniques are often applied. Generally speaking, " RS help agents to find content, products, or services by aggregating and analyzing suggestions from other agents, which means reviews from various authorities and agents " [PKCK12]. Globally, popular RS methods are categorized into two schools: content-based filtering (CB) and collaborative filtering (CF). CB analyzes a set of alternatives rated by an agent and uses the content of the alternatives, sometimes with the provided ratings, to infer an agent's profile that can be used to recommend additional alternatives (items) of interest. In CF, the agent will be recommended alternatives based on other agents with similar tastes and preferences. Besides, hybrid methods combining CB and CF have also been proposed [BS97, MCG+99, Paz99]. Some literature review works providing more detailed introduction are available in [PKCK12] and [AT05].
In addition to RS, preference information is also useful in community detection. In CF, the search of similar agents is actually a process close to the principle of community detection. [JPW07] applied community detection methods on bidding preferences in eBay market data. [YST+13] proposed a RS based on preference-aware community detection. [ZKL15] introduced preference-based non-negative matrix factorization (PNMF) model to incorporate implicit link preference information for overlapping community detection.

4.5 Preference clustering

As mentioned above, in CF, recommended alternatives to an agent are based on other agents with similar tastes and preferences. Such work of finding neighborhood in agents based on the similarity is also also called community detection in domain of

[2]Recommendation Systems are also called recommender systems, or personalization systems in different research communities

social networks, and preference information has already applied in many works such as [LCLM16, YST+13]. With the terminology of preference learning, these processes are clustering on preferences.

The objective is to segment agents into different groups (called clusters) based on their preference information such that agents in the same cluster share more similar preferences to each other than to those in other clusters. This procedure is important in many preference based applications, such as recommendation systems building, community detection, ***etc.***

A brief workflow of data clustering is illustrated in Figure 4.2. The data clustering consists of four principle steps:

1. Feature selection
   In this step, features are extracted and selected to represent every instance (or "agent" in preference clustering), making up the data for further process. In our case, the features are based on preference information of each agents.

2. Clustering algorithm selection and application
   In this step, proper clustering algorithm as well as corresponding parameters are selected and executed on data for processing. After this step, instances are grouped into different clusters.

3. Clustering result validation
   In this step, the clustering results obtained in the previous step are evaluated. The result is validated if the evaluation result is satisfying enough.

4. Interpretation of clustering results
   In this step, the validated clustering results are interpreted as knowledge. Concerning on the clustering on agents, the results can be interpreted as communities of agents.

In this thesis, we mainly focus on the modeling and representation of agents (also called features) from preference data. A major issue particularly important for preferences exists in the completeness of the data. Given an agents set of size K $\mathcal{AGT} = \{agt_1, agt_2, \ldots, agt_K\}$ represented by their partial preferences on alternatives $\mathcal{A}$, clustering based preference elicitation methods consists in two procedures:

1. Cluster agents into different partitions $\mathcal{C} = \{c_1, \ldots, c_{|\mathcal{C}|}\}$ based on partial preference information. Denote the clustering function as $\mathbf{C}(\mathbf{agt}) : \mathcal{AGT} \to \mathcal{C}$.

2. Elicit the unobserved preferences of agent agt_m, based on the partition (or neighborhood) where agt_m belongs.

We can say that this work is a "chicken or the egg causality dilemma". The two procedures are dependent to each other, ***i.e.*** for the agent clustering procedure in an identical algorithm, preferences with less unobserved data returns a more reliable while

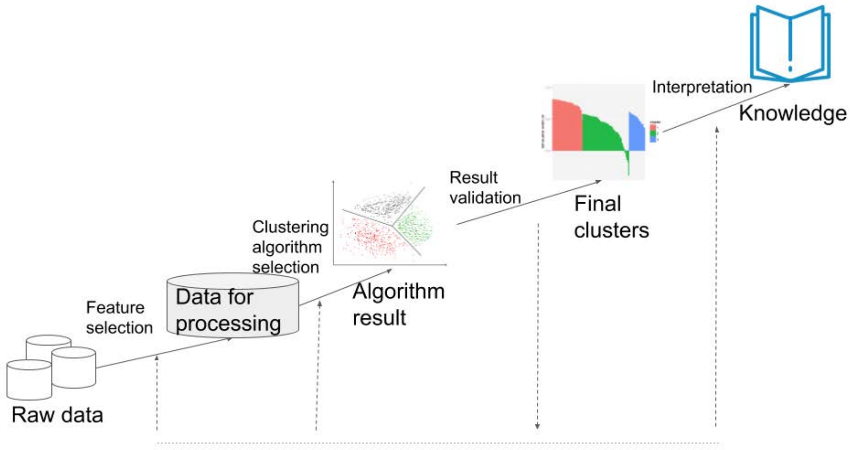


Figure 4.2 – Workflow of data clustering

for the preference elicitation procedure, a partitioning result with better quality returns a more solid prediction.

Hereby, we introduce main stream methods of the clustering on incomplete preferences.

4.5.1 Clustering on incomplete preferences

Clustering on incomplete preferences is a specific application of clustering on data with missing or unobserved information, which is a common issue in machine learning. According to [LR19], the missing data may follow different randomness, divided into three categories, quoted below.

"

1. Missing completely at random (MCAR): This is the highest level of randomness. It occurs when the probability of an instance (case) having a missing value for an attribute does not depend on either the known values or the missing data. In this level of randomness, any missing data treatment method can be applied without risk of introducing statistical bias on the data;
2. Missing at random (MAR): When the probability of an instance having a missing value for an attribute may depend on the known values, but not on the value of the missing data itself;
3. Not missing at random (NMAR): When the probability of an instance having a missing value for an attribute could depend on the value of that attribute.

"

Before the discussion on incompleteness, we firstly distinguish two concepts "sparse matrix" and "scarce matrix".

Definition 4.5.1. Sparse matrix: A sparse matrix (or sparse array) is a matrix in which most elements are zeros.

Definition 4.5.2. Scarce matrix: A scarce matrix is a matrix in which most elements are missing (we call it null in this work).

In many works, the definition of "sparsity" and 'scarcity" are not distinguished. Sparse matrix may refer both of the two scenarios above and missing data are valued as zeros. For matrix on preference information, this is pertinent in following cases:

- The matrix represents the information of comparability. In an information matrix [CKS86] $I = (I_{ij})$, where i, j denotes comparison information between alternatives a_i, a_j, i.e.
$$I_{ij} = \begin{cases} 1, & \text{if } a_i, a_j \text{ are compared,} \\ 0, & otherwise \end{cases} \tag{4.2}$$
Such cases corresponds to MCAR or MAR categories.
- The missing information can be induced as "dislike" opinion. For instance, in a voting mechanism where agents express their top-k alternatives, the missing information on other alternatives can be induced as "dislike". Thus, it is pertinent to define the utility function, such as scores, on these alternatives as zeros.
It should be noted that data are not randomly missing, it's a NMAR case.

Targeted on the scarcity problems where parts of data are null, ***i.e.*** cases of MCAR and MAR, clustering methods can be categorized into three schools:

- **Data imputation**. Data imputation refers to the process of replacing missing data with substitute values. The imputation of missing preferences is already elicitation procedure. Usually, such methods fill the missing data with statistical approaches, such as Neibourhood based methods [BM03, LDSS04], EM based methods [GJ94, WLXC05], Non-negative Matrix Factorization [ZWFM06], ***etc***. The neighborhood based methods imputes the missing data from the nearby agents. Usually, the neighborhood are defined by clustering process (introduced later) or ground truth information. In [GJ94], the authors assume that data instances are generated independently from a mixture probability density and applies EM method to maximize the likelihood on observed values to estimate the parameters of the distribution. More precisely, mixture of Gaussian is applied on real-valued data and mixture of Bernoulli for discrete valued data. In [WLXC05], the authors applied logistic regression model as the probability distribution assumption and applied EM method for the likelihood maximization. Matrix Factorization (MF) is a popular method for score prediction in recommendation systems. Given the scoring matrix from all agents on all alternative $X_{m\times n}$ (n alternatives by m agents), MF aims to find two matrix $U_{m\times k}$ and $V_{k\times n}$ such that

$$X = UV. \tag{4.3}$$

This method assume that the scores follow a linear model, thus the objective is to build a linear model adapted to observed values, usually evaluated also by likelihood. The searching for U and V in MF is also a clustering process where k clusters are determined.

As transitivity is an extra property for preference data comparing to other information, imputation by transitivity of preference relations [ACH$^+$08] (or deduction) is also a reasonable method.

- **Clustering on observed preference (Ignoring and discarding data)**. In this category, all missing data are discarded and the clustering is effected only on observed attributes. There are two mainstream ways of discarding in preference clustering: clustering on top-k preferences or discounting the similarity between objects by weighting on their completeness, also nameed partial distance strategy. In clustering on top-k preferences, alternatives with missing preference information are assumed to be disliked. This assumption transfer the "scarcity" into "sparsity" by filling the "null" with zero. Discounted distances are normally relevant to the superposition of alternatives shared by two partial orders. Many measures have been proposed such as Jaccard similarity [KBGM09], CPCC [SM95], WPCC [HKBR99], SPCC [JE09], ***etc***, corresponding to different measure of similarity between preference orders.

- **Modeling the missing features by soft methods**, such as fuzzy theory [HB01, ZC03] and rough set theory[LDSS05]. These methods interpret the null values with a high degree of uncertainty, then apply soft clustering methods.

The last category is also related to clustering on uncertain data. With uncertainty, data values are no longer atomic. In order to apply traditional clustering methods, uncertainty data needs to be summarized into atomic values. To take mean values is an intuitive way, and such method has been applied on incomplete preference data [Kam03a]. However, taking mean value is based on rough assumptions and could seriously affect the quality of clustering results. Interpreting the uncertain data into interval data is also a feasible solution [VdCL00, CL02]. In fact, such methods are intrinsically based on the assumption that the missing data respect a uniform distribution.
Furthermore, many work apply more sophisticated probability density function (pdf) or statistical models. [HG05] applied maximum likelihood estimation on Mixture Gaussian model. Fuzzy theory have also been applied in clustering on data with uncertainty. In [CCKN06], the authors proposed a fuzzy modeling on uncertain data named Uncertain k-means (UK-means). [NKC$^+$06] studied various prunning methods on UK-means to improve the efficiency. [PLP11] proposed a Uncertain Customized k-means (UCK)

by extended UK-means with the assumption that standard deviation of measurement data is available. For more reading, [CCK$^+$05] provides some synthetic information.
For better understanding, we distinguish two definitions on soft methods hereby. In the domain of soft computing, soft clustering is a popular topic. Soft clustering is different from modeling data with soft methods (soft modeling). In soft clustering, data is crisp. One data point (instance) can belong to multiple clusters with a certainty degree, such as FCM [BEF84] based on fuzzy theory, ECM [MD08], EVCLUS [DM04] and EKNNclus [DKS15a] based on TBF. However, in soft modeling, data are uncertain (say soft data), which is a different scenario than soft clustering. Thus, soft clustering can not be applied directly on soft data.

4.6 Conclusion

Two aspects of preference management are introduced in this chapter: preference aggregation and preference learning (preference clustering). In preference aggregation, we reviewed several typical problems: Condorcet's paradox (transitivity), missing information, as well as Arrow's impossibility theorem. The first two are to avoid or deal with once encountered while Arrow's impossibility describes a property of voting systems.
For crisp preferences, Condorcet's paradox is avoided by the definition of voting mechanism (strategy). For fuzzy preferences, the transitivity is guaranteed by keeping the consistency in fuzzy values, such as additive consistency and multiplicative consistency.
The incompleteness problem is a challenge for both preference aggregation and preference clustering. Depending on various causes of incompleteness, different solutions are applicable, concluded in Table 4.2. For top-k preference, the preference information

Table 4.2 – Different completing strategies to different causes of missing information

Completing strategy	Examples	Preference Aggregation	Preference Learning
As disliked	Top-k are given	$\surd$	$\surd$
As indifference	Absence of agents	$\surd$	x
By statistical model	Average of other agents	x	$\surd$

are given only on the most preferred alternatives. Thus, it's reasonable to induce that the alternatives without preference information given are disliked and such induction is objective.
For agents who are absent in expressing their opinion on preference, it is reasonable to consider their preferences as "indifference", which plays a neutral role in the aggregation process. However, such strategy is no longer applicable in preference learning. The reason is simple: "indifference" is a meaningful value in agent portrait. If such assumption is accepted, all agent not expressing preference opinion will be grouped into one cluster. Such clustering result does not correspond to the initial objective: clustering agents by their portraits of preferences.
For agents who do not express their preference opinion on the entire alternative space,

it is reasonable to complete the missing value by statistic methods, such as average value of preferences from similar agents. However, these induced preferences are not convincing, ***i.e.*** there is no proof that the completed information respects the agent's subjective opinion. Therefore, if such induced preference data are applied in preference aggregation, the aggregated result will lose the reflection to the group's opinion.

In preference learning, a brief plan of problems is introduced. A topic map on preference learning is concluded in Figure 4.3. We also introduced the characteristic

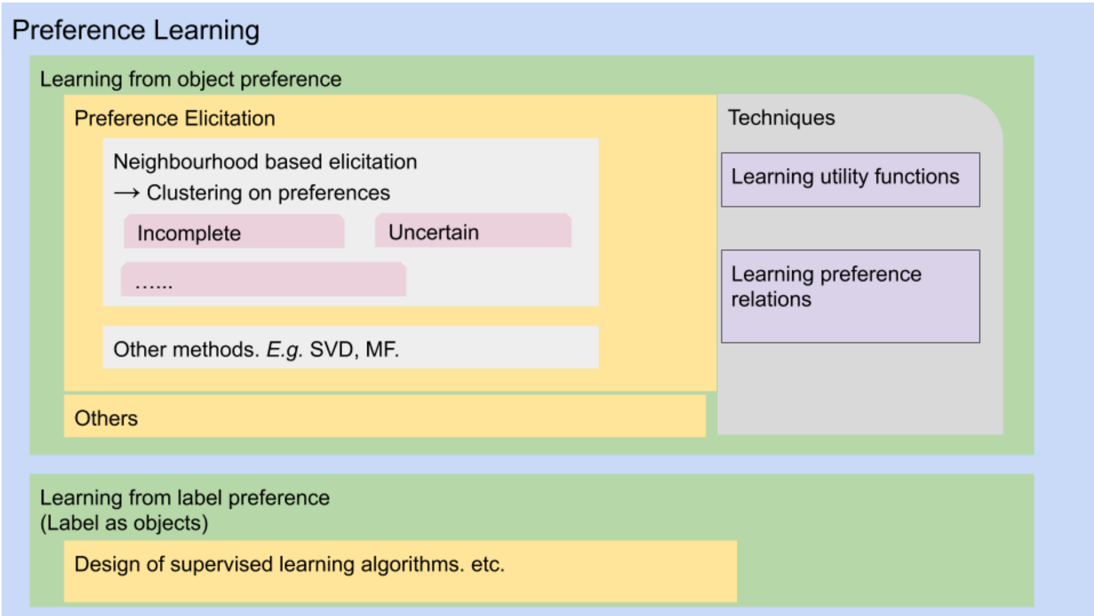


Figure 4.3 – Topic map of preference learning

relation between some aggregation methods and distances. All preference aggregation methods are based on one principle: to find a consensus opinion, i.e. the aggregated preference should have as little conflict to agents' individual preference as possible and respect the majority opinion in agents. The distance based aggregation is equal to the centroid calculation in preference clustering on the same distance for preference similarity measuring.

Some difficulties in learning on imperfect preferences are pointed out, with a brief survey on different schools of clustering methods based on imperfect data, some of which are adaptable for imperfect preferences. In addition to the strategies of completing the preference information, more solutions are applicable for preference learning. A summary is given in Table 4.3.

We also introduced the relation with preference clustering and recommendation systems, which is a major application for preference learning. Finally, with the help of different soft computing methods, such as fuzzy theory and TBF, we distinguished the

Table 4.3 – Methods on clustering on incomplete preferences.

Category	Description	Examples
Complete missing values	Complete missing values under assumptions Usually with zeros or with average value	Table 4.2
Cluster on observed data	Weight the instances with completeness of observation	CPCC WPCC
Soft modeling	Regard the missing data as uncertain problem. Apply soft theories.	UK-means

soft clustering methods and soft modeling for data with uncertainty.

In next part, we introduce contributions of our this thesis work. The contribution is mainly on modeling on imperfect preferences, with applications around aggregation and clustering.

Part III

Contributions

Chapter 5

BFpref model, an evidential model for imperfect preferences

In the introduction chapter, preference relations with imperfectness as well as some ambiguities in the basic definition of preference relations were introduced, identifying new challenges in preference modeling. In Section 2.2 of Chapter 2, it is mentioned that the state of the art models for non crisp preferences are able to modelize the preferences with uncertainty but not imprecision. Besides, in the conclusion part of Chpater 3, an ambiguity in the definition of "incomparability" was introduced. Targeting on these issues, based on the theory of belief functions, we propose a frame of discernment on all possible preference relations between two alternatives, namely BFpref (Belief Function based PREFerence) model [ZBM17]. BFpref model is able to express both uncertainty and imprecision as well as information missing in preference reasoning.

In this chapter, we introduce our contributions on evidential preference modeling. Firstly, we pose an issue on the definition of "incomparability" with different interpretations. Afterwards, a frame of discernment for BFpref model is given in Section 5.2, followed by a solution for modeling the imperfectness, including problems of "uncertainty", "imprecision" and "incompleteness". Afterwards, we propose a strategy to avoid Condorcet's paradox as well as a Depth First Search (DFS) based algorithm to optimize the search of Condorcet's paradox situation.

5.1 An issue in "incomparability"

Incomplete orders (partial orders and quasi orders) include the "incomparability" preference relation. However, "incomparability" relation may refer to multiple interpretations. We compare three examples respectively caused by "absence of knowledge", "conflict" and "non-observed data".

In Chapter I, example 3 presented a situation that agents do not know "Ram-

boutan", so they are not able to give a "strict preference" when alternatives concern "Ramboutan". In this case, "incomparability" is caused by lack of knowledge on the alternatives.
The second example is a real case on Sushi preference dataset [Kam03a], already given in Example 4 in Chapter I. In this example, the "incomparability" is caused by "non-observable" (or implicity) in parts of data. More details of Sushi preference dataset is given in Section 6.3.2.
The third example is taken from [DMÖ+12] between two job offers.

Example 9. Which job is better?

> Bob gets two job opportunities: job_1 and job_2. Job_1 comes with a low salary but is very interesting while job_2 with a high salary but is not interesting at all. When comparing these two jobs with high conflict, Bob has difficulties to express a strict preference in favor of one of them or indifference.

Three scenarios of "incomparability" is demonstrated in the examples. The first two examples are caused by the absence of information (respectively from the angle of agents and the data). The third one is caused by conflicts. In our opinion, different interpretations of "incomparability" should be modeled differently. In next sections, we introduce the BFpref model for uncertain and imprecise preferences, we also distinguish the two interpretations of "incomparability".

5.2 BFpref model

In the BFpref model, we consider a case that a set of agents $\mathcal{AGT} = \{agt_1, \dots, agt_n\}$ expressing their preferences between every pair over an alternative set $\mathcal{A} = \{a_1, \dots, a_k\}$. Thus, each agent is a source of preference information. For any $a_i, a_j \in \mathcal{A}$, four relations are possible:

- Strict preference: $a_i \succ a_j$
- Inverse strict preference: $a_i \prec a_j$
- Indifference: $a_i \approx a_j$
- Incomparability: $a_i \sim a_j$

To ensure that the four relations are exclusive and exhaustive for all preferences between two alternatives, the relation of "incomparability" is indispensable. Therefore, on a given pair of alternatives a_i and a_j, the frame of discernment is defined as:

$$\Omega_{ij} = \{\omega_{ij}^R | R \in \{\succ, \prec, \approx, \sim\}\}. \tag{5.1}$$

Obviously, with the theory of belief functions, BFpref model is able to express preference information with ignorance and imprecision. This model is also available to distinguish different interpretations of "incomparability".
To summarize the three examples of incomparability (Example 3, 4 and 9), given two alternatives $a_i, a_j \in \mathcal{A}$, an "incomparability" relation may refer to two scenarios:

1. a_i and a_j are not compared because of the lack of knowledge or the information of comparability is missing (say "undecided" case);

2. a_i and a_j are not comparable because of conflicts (say "indecisive" case).

To distinguish these two scenarios in BFpref model, the "***indecisive***" case is represented by the singleton $\omega^{\sim}$ and the "***undecided***" case by the vacuous BBA $m_?$ where $m_?(\Omega) = 1$, also representing the total ignorance (see Definition 2.3.4 of Chapter 2).

In some circumstances such as surveys, or voting systems, four BBAs

$$m_s^R, R \in \{\succ, \prec, \approx, \sim\}$$

are defined for agent agt_s, estimating on each alternative pair a_i, a_j. Each BBA is a simple supported, respectively representing the uncertainty degree on the four preference relations and the agents are assumed to be cognitively independent. Such design is for sake of agents' convenience because in a voting system, a person usually feels more convenient to give belief degrees on one single event rather than four. Besides, observation on one relation is always simpler than on four relations. The assumption of cognitive independence is an essential condition required by many combination rules such as conjunctive rule. This assumption is reasonable to be applied on different agents since in many preference collection mechanisms such as voting systems, the exchange between different agents is considered as limited.

In Chapter 4, we introduced preference management as an important research domain. Based on the BFpref model, the management of preferences is studied as well. In this chapter, we mainly focus on the aggregation problems in BFpref and corresponding solutions, introduced in next sections.

5.3 Evidential preference aggregation strategies

As mentioned in Chapter 4, the aggregation of preferences is indeed a specific case of information fusion process, where the theory of belief functions plays an important role. In this section, we mainly introduce two different aggregation rules on evidential preferences based on the combination rule in the theory of belief functions. The entire information fusion process mainly consisting of two steps: combination and decision. Specially for preference data, the avoidance of Condorcet's paradox situations is also an important issue. Thus, we also studied the problem and proposed a Condorcet's paradox avoidance strategy as well as an efficient elimination algorithm.

5.3.1 Combination

The preference aggregation on BFpref model is actually the combination and decision steps with the terminology of the theory of belief functions. In this section, we first introduce different combination strategies for BFpref model, then, in the decision part, we introduce our algorithms to avoid Condorcet's paradox.

Developed from combination and decision rules, we propose two combination strategies, marked as Strategy A and Strategy B. For each alternative pair a_i and a_j $(i < j)$, the strategies are illustrated in Figures 5.1 and 5.2 followed by detailed explanations. Strategy A was originally proposed by [EBMBY15]. Both strategies are based on the two combination rules described in Equations (5.3) and (5.2).

- **Strategy A:**

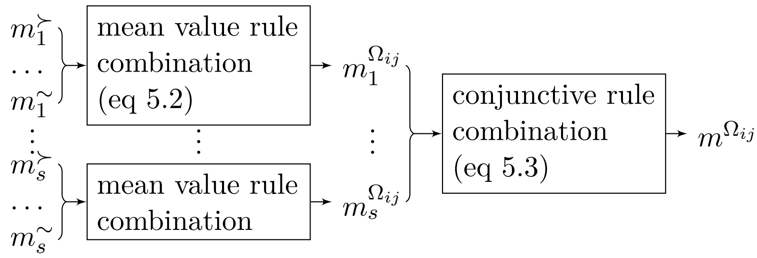


Figure 5.1 – Combination Strategy A

Firstly, we combine the four BBAs on four relations of one pair (a_i, a_j) from one agent agt_s into one BBA. As the four BBAs originate from one identical agent, they are **not** ***cognitively independent*** (see Section 2.3), then, the mean value rule is proper for the combination. Applying mean value combination rule upon equation (5.2) as:

$$m_s^{\Omega_{ij}}(X) = \frac{1}{4} \sum_{R \in \{\succ, \prec, \approx, \sim\}} m_s^R(X), X \subseteq \Omega_{ij} \tag{5.2}$$

where $m_s^{\Omega_{ij}}$ denotes the belief degree of agent agt_s on the entire frame of discernment Ω_{ij} defined in Equation 5.1 . Then, the conjunctive combination rule in equation (5.3) is applied over multiple agents because they are ***cognitively independent***. For an alternative pair (a_i, a_j), we obtain a BBA given by:

$$m^{\Omega_{ij}}(X) = \bigcap_{s=1}^{S} m_s^{\Omega_{ij}}(X), X \subseteq \Omega_{ij} \tag{5.3}$$

where $m^{\Omega_{ij}}$ denotes the aggregated belief level from all agents on the entire frame of discernment. The BBA $m^{\Omega_{ij}}(X)$ is the finally combined BBA for the alternative pair (a_i, a_j). One of the drawbacks of strategy A is in its inability to scale with the volume of information sources (agents in our case). After the first combination with mean value rule, the combined BBAs are no longer simple support BBAs and have auto-conflict. The non-simple-support BBAs cause global-conflicts and lead to increase the value of $m^{\Omega_{ij}}(\emptyset)$ when conjunctive combination rule is applied. Once the volume of sources gets large, global-conflicts are exacerbated by the conjunctive combination rule so that $m^{\Omega_{ij}}(\emptyset)$ converges to 1 [MJO08]. To avoid such deterioration, we proposed strategy B in which the conjunctive rule is applied on simple support BBAs.

- **Strategy B:**

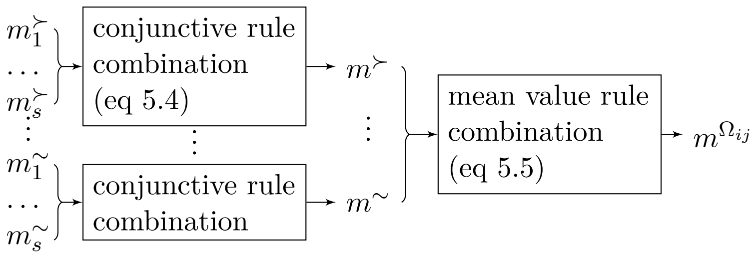


Figure 5.2 – Combination Strategy B

Strategy B is the inverse of strategy A, presented in Figure 5.3.1.
Firstly, we arrange BBAs for each alternative pair (a_i, a_j) of all agents S into 4 groups, respectively representing the four preference relations. On each group, the conjunctive combination (equation (5.4)) is applied.

$$m^{R_{ij}}(X) = \bigcirc\!\!\!\!\cap_{s=1}^{S} m_s^{R_{ij}}(X), R_{ij} \in \{\succ, \prec, \approx, \sim\}, X \subseteq \Omega_{ij} \tag{5.4}$$

where $m^{R_{ij}}$ denotes the aggregated belief degree of all agents on preference relation R between alternatives a_i and a_j. Afterwards, we apply mean value combination method on the 4 combined groups.

$$m^{\Omega_{ij}}(X) = \frac{1}{4} \sum_{R \in \{\succ, \prec, \approx, \sim\}} m^R(X) \tag{5.5}$$

The Strategy B is more suitable when few BBAs are valued with ignorance or imprecision. Applying conjunctive combination rule on such BBAs can cause high global-conflict, ***i.e.*** the combined BBA has large value on $\emptyset$. Such phenomena will be illustrated in the experiments part.

5.3.2 Decision

After two combinations, we finally get a BBA for each pair (a_i, a_j) denoted by $m^{\Omega_{ij}}$. The decision related to the relationship of each pair is taken based on the pignistic probability $betP(\cdot)$ on the space Ω_{ij} by:

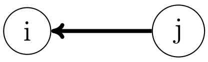
$$\omega_{ij,d} = \underset{\omega_{ij}^R, R \in \{\succ, \prec, \approx, \sim\}}{\operatorname{argmax}} \quad betP^{\Omega_{ij}}(\omega_{ij,R}) \tag{5.6}$$

Since the decision is made on aggregated preferences. Conflict preferences such as Condorcet's paradox may appear. In the subsection 5.3.3, we propose two algorithms to avoid Condorcet's paradox in the final result.

5.3.3 Condorcet's paradox avoidance in graph construction

In the following, we use directed graph for the graphic representation of the preference order obtained from the fusion process. The four possible relationships are illustrated as follows:

- $\omega_{ij}^{\succ}$: $a_i \succ a_j$: 

- $\omega_{ij}^{\prec}$: $a_i \prec a_j$: 

- $\omega_{ij}^{\approx}$: $a_i \approx a_j$: 

- $\omega_{ij}^{\sim}$: $a_i \sim a_j$: 

In the preference graph, a Condorcet's paradox is represented by strongly connected components (or cycle). A strongly connected component of size 2 is considered as an indifference relationship, therefore Condorcet's paradox is represented by cycles of minimum size 3. A simple example of Condorcet's cycle is illustrated in Figure 5.3.

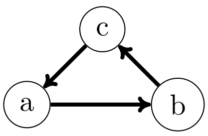


Figure 5.3 – Graphic representation of Condorcet's paradox

Condorcet's paradox does not respect the property of transitivity, which must be satisfied for quasi order (see Definition 2.1.7 in Section 2.1). To avoid Condorcet's paradox, we have to cut an edge in the cycle of size equal or larger than 3. The fact of removing an edge between a and b is equivalent to replace the original relation between a and b by "incomparability". In order to introduce as little knowledge as possible, we decide to remove the edge which is most similar to the relation "incomparability". To measure this dimilarity, in our work, we choose Jousselme distance [JGEB01] (see Equation (3.16) in Chapter 3) for distance measurement. Jousselme distance is considered as a reliable similarity measure between different BBAs [EMSBY14]. It considers coefficients on the elements composed by singletons. In this paper, Jousselme distance is denoted by d_J. Hence distance between an alternative pair (a_i, a_j) and "incomparability" is denoted by $d_J(m_{ij}, \omega_{ij}^{\sim,0})$, where the BBA of incomparability $\omega_{ij}^{\sim,0}$ denotes the categorical BBA ($m(\omega^{\sim}) = 1$) on preference relation $\sim$ between the two alternatives.

Obviously, preferences over multiple alternatives without Condorcet's paradox can be represented by a Directed Acyclic Graph (DAG) (note that cycles of 2 elements are tolerated in such graph because of the "indifference" relation). To simplify the explanation, DAG with tolerance of 2-element-cycle is denoted by DAG_2.

With strongly connected components (SCC) search methods such as Tarjan's algorithm [Tar72], we firstly propose a naive algorithm by iterating "DAG detect-edge remove" process in Algorithm 2. Each iteration detects all the SCC of size larger than 2 as subgraphs (function SCC in Algorithm 2) and remove one edge closest to incomparability in each sub-graph (***Loop Process*** in Algorithm 2).

Algorithm 2 Naive DAG_2 Building Algorithm

Input: A preference graph G constructed on equation (5.6) with edges valued by BBAs $edge.mass$
Output: A directed acyclic graph
  // Loop Process
while $subgraphList{=}SCC(G)$ is not empty **do**
  for *subgraph* in *subgraphList* **do**
    remove *edge* in all *edges* of *subgraph* who is nearest to the preference relation of "incomparability".
  end for
end while
  // SCC search function (apply Tarjan's algorithm)
function SCC(G)
  Input: directed graph G
  Output: The sub-graphs of strongly connected components of size larger than 2 in G.

Since the SCC search function returns the largest strongly connected component found, this method loses its efficiency confronting SCC with nested cycles. A simple structure of nested cycles is illustrated in Figure 5.4.

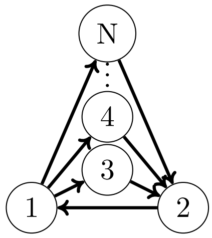


Figure 5.4 – Nested cycles

Facing such structure of preferences, we propose here a more efficient algorithm to build a DAG_2 in an incremental way, described in Algorithm 3. In this algorithm, all edges are ordered by their Jousselme distance to incomparability at the initialization phase (line 2). "Incremental" means that the graph is built by adding edges one by one in a descending order of their distance to incomparability. Given an alternative pair a_i, a_j with their predefined comparable (preference, inverse preference or indifference) relation r calculated from the BBA, we check if the graph is still a DAG_2 with new edge between a_i, a_j added. The checking process is based on a Depth-First-Search (DFS) algorithm. Given two nodes a_i and a_j in graph G, function $DFS(G, a_i, a_j)$ returns the path length from a_i to a_j.

More precisely, if a relation r is a strict preference $a_i \succ a_j$, the DFS algorithm searching node a_i starts from node a_j. If a_i is found, a cycle will appear if the edge $i \longrightarrow j$ is added. In such case, we replace the relation between a_i, a_j by an incomparability relation (remove the edge) (line 6 and 7). Similarly if r is inverse preference $a_i \prec a_j$, the DFS algorithm searches node a_j from a_i (line 8 and 9). However, the relation “indifference" may hinder the time performance of this algorithm. If r represents the indifference relation $a_i \approx a_j$, we have to apply DFS twice from a_i to a_j and from a_j to a_i (line 10 and 11).

In a structure of nested cycles containing E edges, V vertex and N cycles, the naive algorithm based on SCC search (Algorithm 2) can reach a temporal complexity of $\mathcal{O}(N(|E| + |V|))$ while the incremental algorithm (Algorithm 3) has a temporal complexity of $\mathcal{O}(N)$.

Although the incremental algorithm is efficient on nested cycle, its time performance degenerates when the preference structure has few nested cycles or many indifference relations. The applicability of the two algorithms is demonstrated and discussed in the following section.

5.4 Experiments

In this section, we compare the two proposed preference fusion strategies and two proposed algorithms for Condorcet’s paradox avoidance. The fusion strategies are evaluated from a numeric point of view while the algorithms are evaluated in terms of time performance. Lacking social network data from real world, the data used in our experiments were generated manually or randomly.

5.4.1 Preference fusion strategies

In this experiment, we firstly define preference structures of 3 agents as illustrated in figure 5.5. To simplify our experiments, the alternative pairs are always in an ascending order (***i.e.*** $\forall a_i, a_j \in A \Rightarrow i < j$). The belief degrees for alternative pairs (2,3), (2,4) and (3,4) are specially given in table 5.1. For the other alternatives, their belief degrees are set by default values given in table 5.2.

Algorithm 3 Incremental DAG_2 Building Algorithm

Input: All pairs $PAIRS$ and their BBAs M
Output: A DAG_2 graph
// *Initialization*:
Initialize an empty graph G
Ascending order all pairs $PAIRS$ by $d_J(m^{\Omega_{ij}}, \omega_4^0)$, stock in a stack denoted as $Stack$.
// *Loop Process*
while $Stack$ is not empty **do**
  pair $(a_i, a_j) = Stack.pop()$
  Add nodes a_i and a_j into graph G
  if $a_i \succ a_j$ **then**
    pathLength=$DFS(G, a_j, a_i)$
  else if $a_i \prec a_j$ **then**
    pathLength=$DFS(G, a_i, a_j)$
  else if $a_i \approx a_j$ **then**
    pathLength=max($DFS(G, a_j, a_i)$, $DFS(G, a_i, a_j)$)
  end if
  if pathLength$\geqslant 2$ **then**
    Consider relation between (a_i, a_j) as incomparability
  else
    add edge between (a_i, a_j) calculated by m_{ij} (equation (5.6))
  end if
end while
// *Recursive Function*
function DFS(G, v, n):
label v as discovered
if all successors of v in G are labeled as discovered **then**
  return 0
else
  for w in non-discovered successors of v **do**
    if w is n **then**
      return 1
    else
      len=DFS(G, w, n)
      if $len == 0$ **then**
        return 0
      else
        return $len + 1$
      end if
    end if
  end for
end if

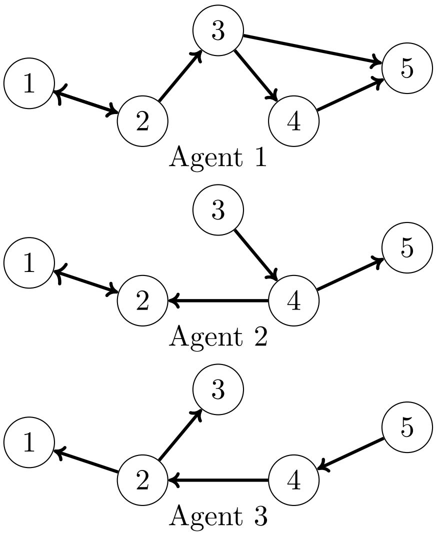


Figure 5.5 – Preference order of three agents

Table 5.1 – Belief degree values for alternative pairs (2,3), (2,4) and (3,4)

agent/pair	ω_1	ω_2	ω_3	ω_4
Agent 1/(2,3)	0.8	0.7	0.6	0.5
Agent 1/(2,4)	0.4	0.1	0.3	0.6
Agent 1/(3,4)	0.9	0.8	0.7	0.6
Agent 2/(2,3)	0.5	0.4	0.6	0.9
Agent 2/(2,4)	0.2	0.4	0.3	0.1
Agent 2/(3,4)	0.9	0.8	0.1	0.7
Agent 3/(2,3)	0.6	0.2	0.4	0.1
Agent 3/(2,4)	0.3	0.5	0.2	0.1
Agent 3/(3,4)	0.8	0.1	0.6	0.9

Table 5.2 – Default belief degree values on 4 relations

relation	ω_1	ω_2	ω_3	ω_4
preference	0.8	0.2	0.3	0.1
inverse preference	0.1	0.9	0.2	0.1
indifference	0.3	0.3	0.7	0
incomparability	0.1	0.1	0	0.9

With both fusion strategies A and B (in Section 5.3), we get two different results illustrated in Figures 5.6 and 5.7.

A different edge result between alternatives a_1 and a_2 is high lighted by dashed dotted line: the relation between alternatives a_1 and a_2 is indifferent from strategy A,

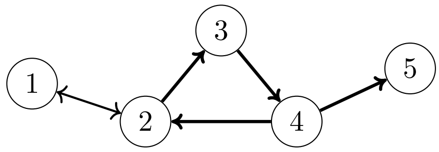


Figure 5.6 – Fusion result of strategy A

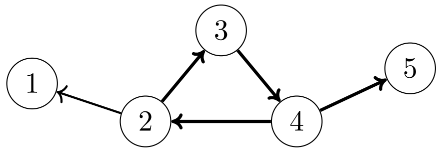


Figure 5.7 – Fusion result of strategy B

while $a_2 \succ a_1$ from strategy B. Focal elements in BBAs on edges are shown in Tables 5.3 and 5.4. Since in original data, BBAs are zero on all non-singleton elements, the final BBAs are still zero on union elements except ignorance (and empty set for strategy A).

Table 5.3 – BBA values on edges of strategy A

pair	$\emptyset$	ω_1	ω_2	ω_3	ω_4	Ω
(1,2)	0.17989	0.08620	0.19870	0.21626	0.01139	0.30755
(2,3)	0.07858	0.17209	0.03612	0.08545	0.15812	0.46962
(2,4)	0.08745	0.07225	0.12931	0.11762	0.12373	0.46962
(3,4)	0.07858	0.17209	0.03612	0.08545	0.15812	0.46962
(4,5)	0.20880	0.22056	0.15581	0.09589	0.03375	0.28519

Table 5.4 – BBA values on edges of strategy B

pair	$\emptyset$	ω_1	ω_2	ω_3	ω_4	Ω
(1,2)	0	0.13975	0.23775	0.232	0.025	0.3655
(2,3)	0	0.1765	0.05	0.10825	0.17	0.49525
(2,4)	0	0.1	0.152	0.138	0.14875	0.46125
(3,4)	0	0.1765	0.05	0.10825	0.17	0.49525
(4,5)	0	0.241	0.234	0.152	0.06775	0.30525

From this result, we observe that:

1. Both combination strategies do not always return same results.
2. A significant difference between the two results exists in the value of empty set $\emptyset$.
3. A Condorcet's paradox appears in the fusion result.

The first two observations concern the empty set value $m^{\Omega_{ij}}(\emptyset)$. If high global-conflict exist in among sources, a conjunctive combination can cause high value on

$m^{\Omega_{ij}}(\emptyset)$, implying high auto-conflict in the combined result [MJO08]. Moreover, this conflict accumulate when the number of sources increasing. As a result, $m^{\Omega_{ij}}(\emptyset)$ may converge to 1 and other focal elements are no longer convincing. For better demonstration, we illustrate the average of $m^{\Omega_{ij}}(\emptyset)$ after combination on all alternative pairs. The average value is calculated by:

$$m_{average}(\emptyset) = \frac{1}{|C^2_{|\mathcal{A}|}|} \sum_{i,j \in C^2_{|\mathcal{A}|}} m^{\Omega_{ij}}(\emptyset) \tag{5.7}$$

where C denotes the combination operation, $|\mathcal{A}|$ the size of the alternative set $\mathcal{A}$.

The agent number ranges from 3 to 90 with step of 3 by replicating Agent 1, 2 and 3 by multiple times. The result is illustrated in Figure 5.8.

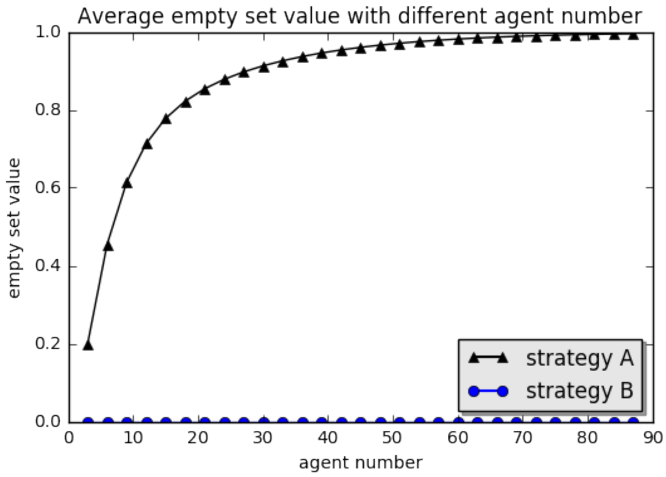


Figure 5.8 – Average of $m(\emptyset)$ with agent number increasing

This result shows clearly the default of strategy A. The average of $m^{\Omega_{ij}}(\emptyset)$ in strategy A converges to 1 while it is always 0 in strategy B. Our experiment example is already an optimal case because agents belief degrees are not various enough (they are just a repetition of the belief degrees of Agent 1, 2 and 3). In fact, the curve of strategy A may rise faster if agents' belief degrees are more various.

In Condorcet's paradox made up by alternatives 2, 3 and 4, the distance between the final combined BBA associated to the edges of the cycle and the relation of incomparability is given in Table 5.5.

Table 5.5 – Jousselme distance between alternative pair BBA and incomparability

alternative pair	d_J in Strategy A	d_J in Strategy B
(2,3)	0.70798	0.60715
(3,4)	0.77016	0.60739
(2,4)	0.66928	0.64387

From table 5.5, another difference between strategy A and B can be found. In strategy A, BBA of alternative pair $(2, 4)$ is the closest to incomparability while in strategy B, the closest alternative pair to incomparability is $(2, 3)$. The final result is illustrated in Figures 5.9 and 5.10.

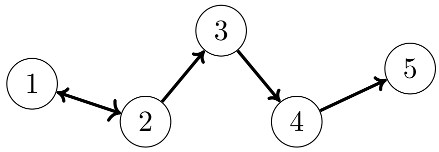


Figure 5.9 – Result without Condorcet's paradox in strategy A

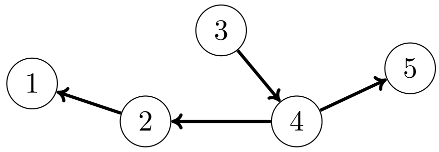


Figure 5.10 – Result without Condorcet's paradox in strategy B

Considering the drawback caused by convergence of empty set value, we believe that strategy B outperforms strategy A, because the latter one does not scale with the number of agents.

5.4.2 Condorcet's paradox avoidance

In this experiment, we evaluated the performance on three special preference structures: nested cycles (Figure 5.4), entangled circles (Figure 5.11) and non-nested structures with indifference relations (Figure 5.12). The BBAs associated to the preference relations are randomly generated adapting with entangled Condorcet's paradox. The evaluation is based on the runtime with increasing number of nested cycles N. As all BBAs are randomly generated following an uniform distribution from 0 to 1, we take the average value of 10 same tests to ensure the reliability of the result. In the experiments, alternative numbers range from 20 to 400 with an interval of 40.

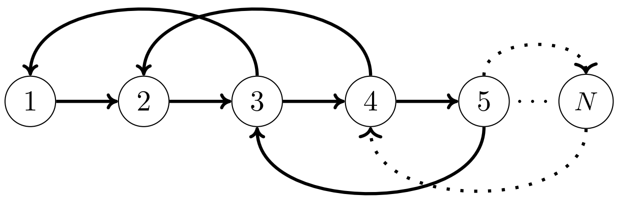


Figure 5.11 – Entangled cycles

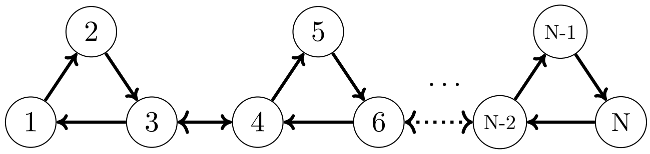


Figure 5.12 – Non-nested Cycles

The performances related to the three preference structures are illustrated in figure 5.13. We observe that the incremental algorithm outperforms the naive algorithm when the structures contain a great number of nested cycles. However, for the structures with little nested cycles but many indifference relations, the naive algorithm performs better. Hence, the selection of Condorcet's paradox avoidance algorithm should be adapted to the structure of the preferences.

5.5 Conclusion

In this chapter, we introduced the modeling and fusion problem of preference with uncertainty and imprecision. Targeted on such preferences information, we proposed and compared two strategies with conjunctive combination rule and mean value combination rule, based on the theory of belief functions. The one which avoid global conflicts performs better when the number of sources scales up. We also proposed a Condorcet's paradox avoidance method as well as an efficient DFS-based algorithm adapting to preference structure with nested cycles. By comparing the time performance of Condorcet's paradox avoidance algorithms on different types of preference structures, we noticed that the incremental algorithm is more efficient on nested structures while the naive algorithm is better on non-nested ones. Limited by our data sources, our experimental works were done on synthetic data. Furthermore, the algorithm for DAG construction can be applied in more general cases, other than those related to preference orders. In domains concerning directed graphs with valued edges (***e.g.*** telecommunication, social network analysis, ***etc.***), Algorithm 3 may find its usefulness.

There are still more works left to explore. Since our experiments are executed on synthetic data, the quality of aggregation result is not objectively evaluated. Concerning the "transitivity" property, only the Condorcet's paradox is removed, but the relation of "incomparability" does not guarantee the property of "transitivity". Another issue falls on the similarity measure applied in decision making. In this decision strategy, the preference of strict preference is equally weighted as indifference, which is in contradiction to the natural sense. When aggregating two preference relations: $\succ$ and $\approx$, it is natural to reach to a $\succ$ relation as result. However, the aggregation method proposed in this chapter gives equal chance to these two relations. This issue will be discussed in detail and solved in the next chapter.

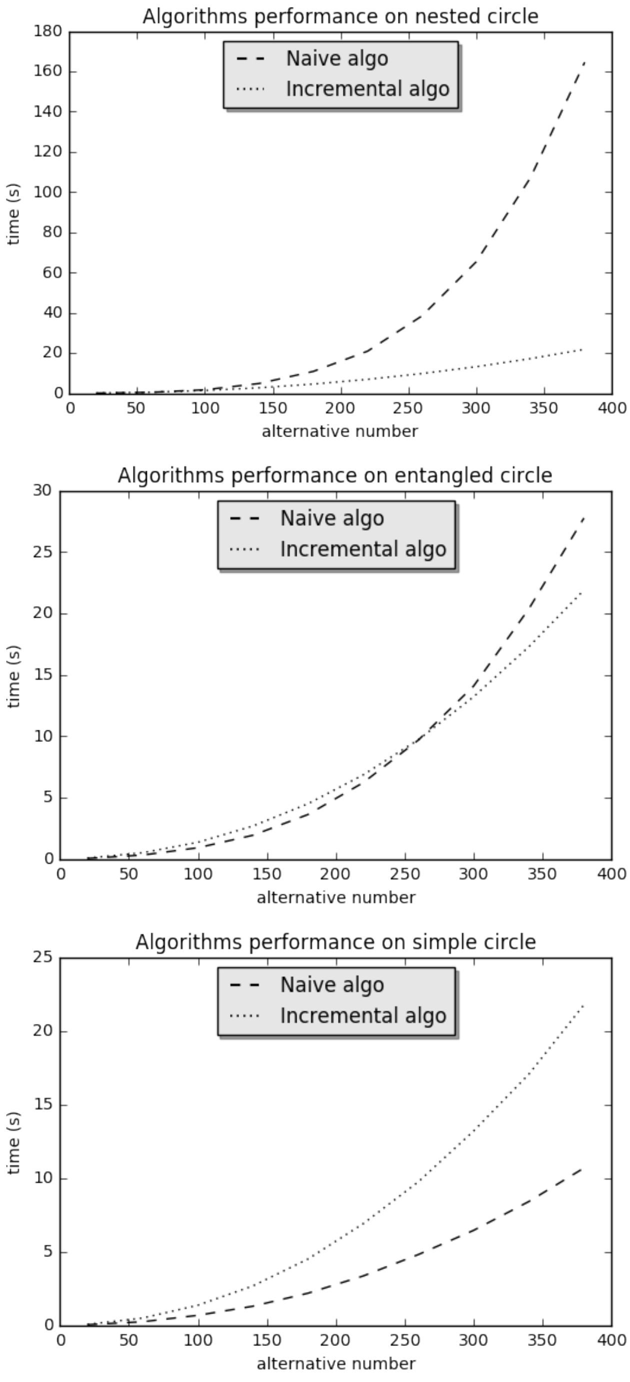


Figure 5.13 – Performances on different preference structures

Chapter 6

A weighted singleton distance for BFpref model

In the previous chapter, a belief function based model for imperfect preference was proposed. At the end of the last chapter, we mentioned an issue in the distance for BBAs that the difference between singletons are always equal, which may be against the properties of the singletons in some circumstances. In decision making process, this issue may cause a biased result when confronting the relation of "indifference". In this chapter, we introduce our contribution on a novel distance for BBAs while considering different weights on singletons, named Weighted Singleton Distance (WSD). The advantage of WSD exists mainly in decision making on evidential preferences (BFpref model). This chapter is structured as follows. In Section 6.1, we firstly review the assumptions in the properties in the state-of-the-art distances on BBAs and then attribute the issues to a redundant property. In Section 6.2, by respecting new properties, WSD is introduced by extending on axiomatic distance for preference relation types. In Section 6.3, WSD is applied in preference aggregation and compared with other consensus strategies on both synthetic data and SUSHI data from the real world.

6.1 An issue in the similarity between evidential preferences

In the theory of belief functions, various divergence and similarity measure methods have been introduced for BBAs in Chapter 3. The definition of the frame of discernment $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$ requires all singleton elements to be mutually exclusive and exhaustive, which are essential properties for the modeling in the theory of belief functions. However, most of these metrics accept an assumption that singleton elements are equally measured when measuring the similarity between different singleton elements. Precisely:

Assumption 1. In a set of exclusive elements $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$, the similarity between each elements is constant (normalized as 0). Formally,

$$sim(\omega_m, \omega_n) = 0, \forall \omega_m, \omega_n \in \Omega. \tag{6.1}$$

Apparently, Assumption 1 is inappropriate for BFpref model. As introduced in Section 3.2 of Chapter 3, we observe a contradictory case that $d_\Delta(\succ, \prec) > d_\Delta(\succ, \approx)$, where d_Δ denotes the axiomatic distance between preference relations. For similarity on BFpref model, the following condition should be respected:

$$sim(\omega^{\succ}, \omega^{\prec}) < sim(\omega^{\succ}, \omega^{\approx}) \tag{6.2}$$

where $\omega^R, R \in \{\succ, \prec, \approx, \sim\}\}$ represent the singletons on corresponding preference relations. This inequality relation indicates that the relation $\succ$ and $\prec$ are most incoherent among all couples of relations. The contradiction between Equation (6.2) and (6.1) often appears in decision making step. Even in some distance independent decision making strategies, such as maximum pignistic strategy (see Equation (2.30) in Section 2.3). Indeed, on decision on singletons, maximum pignistic strategy is equivalent to the minimum Jousselme distance (see Equation (3.16) in Section 3.4) to categorical BBAs. A proposition is therefore given:

Proposition 1. A decision is made on BBA m in the frame of discernment $\Omega = \{\omega_1, \ldots, \omega_k\}$, the following rules return the same result.
Given

$$\begin{aligned} \omega_{betP} &= \underset{\omega_i \in \Omega}{argmax}(betP_m(\omega_i)) \\ \omega_{minD} &= \underset{\omega_i \in \Omega}{argmin}(d_J(m, \omega_i^0)) \end{aligned} \tag{6.3}$$

with ω_i^0 denoting a categorical BBA on ω_i, and d_J Jousselme distance [JGEB01], the equality relation

$$\omega_{betP} = \omega_{minD}$$

is always true.

The proposition is thus given in another way:

Proposition . In the frame of discernment $\Omega = \{\omega_1, \ldots, \omega_k\}$. If for a BBA m, exist ω_d such that $\forall \omega_i \in \Omega, \omega_i \neq \omega_d$,

$$betP_m(\omega_d) \geqslant betP_m(\omega_i) \tag{6.4}$$

Then the following in-equation is true:

$$d_J(m, \omega_d^0) \leqslant d_J(m, \omega_i^0) \tag{6.5}$$

where ω_d^0 and ω_i^0 represent categorical BBA on ω_d and ω_i respectively.

The demonstration is given in Appendix A.

With Assumption 1 abandoned, we hereby recapitulate more properties required in similarity measuring between evidential preference relations.

6.1.1 Properties of similarity measure for evidential preferences

The relation between different order types from the most specific to the most general is illustrated in Figure 6.1.

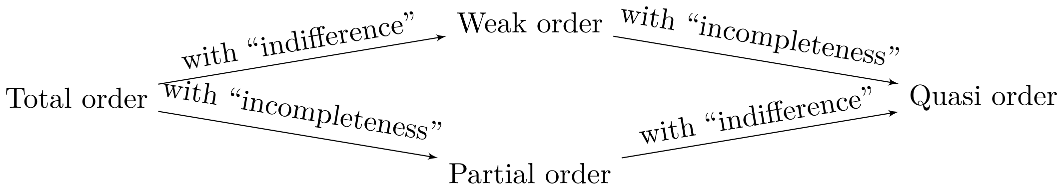


Figure 6.1 – Relations between four types of orders

In this relation map, between any two types of orders, the relation on the right is more general (possessing more relations) than the left, Thus, similarity measure for quasi (pre)orders is compatible with all the other three orders. However, most of the similarity measures are designed for complete orders, ***i.e.*** weak orders and total orders. The difficult part falls on the similarity measure when dealing with "incomparable" relation. To give up Assumption 1 (the similarities between singletons are equal), we define three necessary properties for similarity measure between two BBAs in BFpref model.

Property 1. Weighted singleton: The similarity measure between BBAs takes into account the different similarity between singletons. Scilicet, Assumption 1 in Section 6.1 is abandoned.

Property 2. Metric: The similarity is measured by a distance. Therefore, the properties of metric are kept.

Thus, non-negativity, identity of indiscernible, symmetry and triangle-inequality are kept. These properties guarantee that the similarity is measured by a metric. In the theory of belief functions, BBAs are defined in space of 2^Ω. Hereby, we introduce the property of strongly structural for distances between BBAs. The definition is borrowed from [JM12].

Definition 6.1.1. Strongly structural: A similarity measure *sim* between two BBAs m_1 and m_2 is called *strongly structural* if its definition accounts for the interaction between the focal elements of m_1 and m_2.

We directly give an example of strongly structural property for evidential preferences in Equation (6.6). Hence, the similarity measure for BBA should have stongly structural property.

Property 3. Strongly structural: Formally, a similarity measure is strong structural, if:

$$\forall X, Y \in 2^\Omega, sim(X,Y) \geq sim(X \setminus (X \cap Y), Y). \tag{6.6}$$

For a weak order with $\Omega_{weak} = \{\omega_{\succ}, \omega_{\prec}, \omega_{\approx}\}$, Equation (6.6) requires that the following three relations must be satisfied.

1. Similarity between strict preference and union with indifference:

$$sim(\{\omega_{\succ}\}, \{\omega_{\succ}, \omega_{\approx}\}) > sim(\{\omega_{\succ}\}, \{\omega_{\approx}\}) \\ \Rightarrow sim(\{\omega_{\succ}\}, \{\omega_{\succ}, \omega_{\approx}\}) \in [1-p, 1]. \tag{6.7a}$$

   Symmetrically,

$$sim(\{\omega_{\prec}\}, \{\omega_{\prec}, \omega_{\approx}\}) > sim(\{\omega_{\prec}\}, \{\omega_{\approx}\}) \\ \Rightarrow sim(\{\omega_{\prec}\}, \{\omega_{\prec}, \omega_{\approx}\}) \in [1-p, 1]. \tag{6.7b}$$

   where $1-p$ is the normalized similarity value between $\omega_{\succ}$ and $\omega_{\approx}$

2. Similarity between strict preference and total ignorance:

$$sim(\{\omega_{\succ}\}, \Omega) < sim(\{\omega_{\succ}\}, \{\omega_{\succ}, \omega_{\approx}\}). \tag{6.8a}$$

   Symmetrically,

$$sim(\{\omega_{\prec}\}, \Omega) < sim(\{\omega_{\prec}\}, \{\omega_{\prec}, \omega_{\approx}\}). \tag{6.8b}$$

3. Strict preference and union of other two different preference types:

$$sim(\{\omega_{\succ}\}, \{\omega_{\approx}\}) > sim(\{\omega_{\succ}\}, \{\omega_{\prec}, \omega_{\approx}\}) > sim(\{\omega_{\succ}\}, \{\omega_{\prec}\}). \tag{6.9a}$$

   Symmetrically,

$$sim(\{\omega_{\prec}\}, \{\omega_{\approx}\}) > sim(\{\omega_{\prec}\}, \{\omega_{\succ}, \omega_{\approx}\}) > sim(\{\omega_{\succ}\}, \{\omega_{\prec}\}). \tag{6.9b}$$

In Section 6.2, we introduce an adapted distance measure for BFpref model satisfying all these three properties.

6.2 WSD – a distance for weighted singletons in the discernment of preference types

Jousselme distance is widely accepted in applications based on the theory of belief functions. It respects Properties 2 and 3 in Section 6.1.1. Thus, inspired by the methology of Jousselme distance given in Equation (3.16), we develop the WSD distance with Property 1 respected.

By applying Jaccard index, Jousselme distance satisfies Assumption 1. To address this contradiction, we define WSD distance as follows:

$$d_{WSD}(m_1, m_2) = \sqrt{(\boldsymbol{m_1} - \boldsymbol{m_2})^T \boldsymbol{Sim}(\boldsymbol{m_1} - \boldsymbol{m_2})} \tag{6.10}$$

where $\boldsymbol{Sim}$ is the matrix of similarities between different elements. The calculation of $\boldsymbol{Sim}$ is detailed in the rest of this subsection.

Firstly, we define the similarity function $Sim(X_1, X_2)$ between two elements X_1 and X_2 by two functions $resemb(X_1, X_2, \dots, X_K)$ and $entire(X_1, X_2, \dots, X_K)$:

$$sim(X_1, X_2) = \frac{resemb(X_1, X_2)}{entire(X_1, X_2)} \tag{6.11}$$

where $resemb(X_1, X_2)$ describes the cause of the similarity between X_1 and X_2 (say resemblance), and $entire(X_1, X_2)$ the entire part concerned by X_1 and X_2.
The calculation of $resemb(X_1, X_2, \dots, X_K)$ and $entire(X_1, X_2, \dots, X_K)$ is introduced as follows.

Denote the function $resemb(\cdot)$ and $entire(\cdot)$ with Assumption 1 accepted as $resemb_{ass1}(\cdot)$ and $entire_{ass1}(\cdot)$, the functions $resemb_{ass1}(X_1, X_2, \dots, X_K)$ and $entire_{ass1}(X_1, X_2, \dots, X_K)$ are inherited as:

$$resemb_{ass1}(X_1, X_2, \dots, X_K) = |X_1 \cap X_2 \cap \dots \cap X_K| \tag{6.12}$$

$$entire_{ass1}(X_1, X_2, \dots, X_K) = |X_1 \cup X_2 \cup \dots \cup X_K| \tag{6.13}$$

To simplify the expression, we denote $W = \{X_1, X_2, \dots, X_K\}$, $resemb(W)$ for $resemb(X_1, X_2, \dots, X_K)$ and $entire(W)$ for $entire(X_1, X_2, \dots, X_K)$. Thus, by dropping Assumption 1, $entire(W)$ is defined as a generalized version of cardinal function on union:

$$\begin{aligned} entire(W) \;=\;& \sum_{\omega \in \cup W} entire(\omega) - \sum_{W_2 \subseteq W, |W_2|=2} resemb(W_2) \\ &+ \sum_{W_3 \subseteq W, |W_3|=3} resemb(W_3) - \sum_{W_4 \subseteq W, |W_4|=4} resemb(W_4) \\ &+ \dots + \sum_{W_k \subseteq W, |W_{|\Omega|}|=|\Omega|} resemb(W_k) \times (-1)^{|\Omega|} \\ =\;& \sum_{\omega \in \cup W} entire(\omega) + \sum_{k=1}^{|\Omega|} \sum_{W_k \subseteq W, |W_k|=k} resemb(W_k) \times (-1)^k, \end{aligned} \tag{6.14}$$

where $|W_k|$ denote the number of arguments ($X \subseteq \Omega$) in W_k and $\cup W$ the union of elements in all arguments, ***i.e.*** for $W = \{X_1, X_2, \dots, X_K\}$, we have:

$$\cup W = X_1 \cup X_2 \cup \dots \cup X_K. \tag{6.15}$$

With Equation (6.14), given only the similarity between two singleton elements, $entire(W)$ does not have a unique solution. To guarantee the uniqueness in the solution for similarity, we introduce two other assumptions.

Assumption 2. In a set of exclusive elements $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$, any resemblance is shared only by maximal two exclusive elements, *i.e.*

$$resemb(W) = 0, \quad \forall W \subseteq 2^{\Omega}, |W| \geq 3 \tag{6.16}$$

Assumption 2 guarantees that given only $sim(\omega_m, \omega_n)$, a unique solution exists for $resemb(\omega_m, \omega_n)$.

Assumption 3. Elements are normalized in the calculation. Entire part concerned by one singleton is assigned to 1, formally:

$$entire(\omega) = 1, \quad \forall \omega \in \Omega \tag{6.17}$$

With Assumptions 2 and 3, given similarity $sim(\omega_m, \omega_n)$, the resemblance is calculable. Deduced from Equation (6.14), we have:

$$entire(X, Y) = \sum_{\omega \in X \cup Y} entire(\omega) - \sum_{\substack{\omega_m \in X \\ \omega_n \in Y \\ m \neq n}} resemb(\omega_m, \omega_n) \tag{6.18}$$

Hence, Equation (6.11) is inherited as:

$$sim(X_1, X_2) = \frac{\sum\limits_{\substack{\omega_m \in X_1 \\ \omega_n \in X_2 \\ m \neq n}} resemb(\omega_m, \omega_n)}{\sum\limits_{\omega \in X_1 \cup X_2} entire(\omega) - \sum\limits_{\substack{\omega_m \in X_1 \\ \omega_n \in X_2 \\ m \neq n}} resemb(\omega_m, \omega_n)} \tag{6.19}$$

To guarantee Assumption 2 given by Equation (6.16), the following equation must be respected:

$$\sum_{\substack{\omega_m, \omega_n \in \Omega \\ \omega_m \neq \omega_n}} sim(\omega_m, \omega_n) \leq 1 \tag{6.20}$$

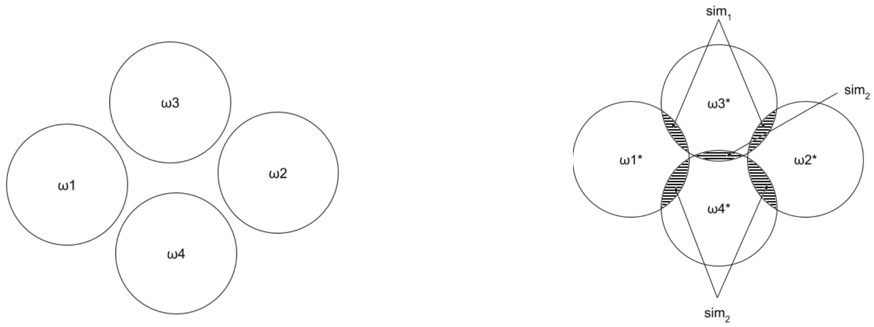
6.2.1 Graphic demonstration of WSD calculation

For better comprehension, we illustrate the calculation of WSD in a graphical way on four singletons $\omega_1, \omega_2, \omega_3$ and ω_4 in Figure 6.2.

According to the exclusiveness of discernment framework, singletons has no superposed part, illustrated in Figure 6.2a. Obviously, Figure 6.2a is isomorphic with Figure 2.1. In Jousselme distance, distances between singletons are equal as no superposed area is shared between any two singletons.

We assume that exclusive singletons may have superposed part in another space S^* different from their original definition space S, causing the difference in similarities between singletons. Such hypothesis is reasonable. For instance, among three cars

Volkswagen Golf, Volkswagen Polo, and Renault Clio, each car is exclusive with others if the description of the car is only on the entity (space S). However, Golf and Polo can be considered as more similar than Clio if the brand of the car is considered for the description (space S^*). Without a concrete definition, the superposed surface is assigned by similarity values between singletons, as illustrated in Figure 6.2b.



(a) Element in original space S

(b) Element in space S^* with different similarity values

Figure 6.2 – Graphical representation of WSD calculation

Thus, the calculation of similarity between elements $X, Y \in 2^\Omega$ becomes the calculation of the corresponding area in Venn diagram of S^* space. Assumption 2 indicates that superposition only exists between at most two singletons while Assumption 3 indicates that the area of each singleton (circle) is valued as 1.

An illustrative example for WSD dissimilarity calculation is given in Section 6.2.2.

6.2.2 Illustrative example-WSD for BFpref model

As aforementioned, four preference relations are exclusive with each other, as shown in Figure 3.4a. For two BBAs m_1 and m_2 on Ω^{pref} (defined in Equation (5.1)), with Assumption 1, in Jousselme distance, similarities between elements are constant, which is contradictory with Equation (6.2) in Section 6.1. Thus, WSD is more appropriate. Hereby, given the similarities between singletons, we illustrate how WSD is calculated as a tutorial. The tutorial takes the example of BFpref model.

The similarity values between $\omega_R | R \in \{\succ, \prec, \approx, \sim\}$ is given in Table 6.1, In Table 6.1, x, y and z represent the values of corresponding similarities between singletons.

Table 6.1 – Similarity between singletons

	$\omega_\succ$	$\omega_\prec$	$\omega_\approx$	$\omega_\sim$
$\omega_\succ$	1	0	x	y
$\omega_\prec$	0	1	x	y
$\omega_\approx$	x	x	1	z
$\omega_\sim$	y	y	z	1

These values are not definitively assigned because various versions exist, as introduced in Section 3.2 of Chapter 3. To simplify the deduction, function $resemb(X_1, X_2)$ is assigned by v_1, v_2, v_3 as follows:

$$\begin{aligned} resemb(\omega_\succ, \omega_\approx) = resemb(\omega_\prec, \omega_\approx) &= v_1 \\ resemb(\omega_\succ, \omega_\sim) = resemb(\omega_\prec, \omega_\sim) &= v_2 \\ resemb(\omega_\approx, \omega_\sim) = \qquad\qquad\qquad & \quad v_3 \end{aligned} \tag{6.21}$$

From Equations (6.17), (6.18) and values in Equation (6.21), following relations are established:

$$\begin{cases} x = \dfrac{v_1}{2 - v_1} & \Rightarrow\ v_1 = \dfrac{2x}{1+x} \\ y = \dfrac{v_2}{2 - v_2} & \Rightarrow\ v_2 = \dfrac{2y}{1+y} \\ z = \dfrac{v_3}{2 - v_3} & \Rightarrow\ v_3 = \dfrac{2z}{1+z} \end{cases} \tag{6.22}$$

where x, y and z are similarity values of Table 6.1. Therefore, with numeric values of v_1, v_2 and v_3, by applying Equation (6.19), the similarity in space 2^Ω can be calculated. To guarantee Assumption 2, we have

$$\begin{aligned} 2v_1 + v_3 \le 1 \quad &\Rightarrow \quad \frac{4x}{1+x} + \frac{2z}{1+z} \le 1 \quad \Rightarrow \quad \frac{2}{x+1} + \frac{1}{z+1} \le \frac{5}{2} \\ 2v_2 + v_3 \le 1 \quad &\Rightarrow \quad \frac{4y}{1+y} + \frac{2z}{1+z} \le 1 \quad \Rightarrow \quad \frac{2}{y+1} + \frac{1}{z+1} \le \frac{5}{2} \\ v_1 + v_2 \le 1 \quad &\Rightarrow \quad \frac{2x}{1+x} + \frac{2y}{1+y} \le 1 \quad \Rightarrow \quad \frac{1}{x+1} + \frac{1}{y+1} \le \frac{3}{2} \end{aligned} \tag{6.23}$$

In Fagin's distance (see Equation (3.14) of Chapter 3), the similarity between "indifference" and "strict preference" is measured by a penalty value p, where $sim(\omega_\succ, \omega_\approx) = 1 - p$. Recall that in Fagin's distance, in order to guarantee the property of metric, the value of p should be chosen by Equation (3.13), repeated here.

$$\frac{1}{2} < p < 1$$

Thus, the similarity is chosen by:

$$0 < sim(\omega_\succ, \omega_\approx) < \frac{1}{2} \tag{6.24}$$

This is consistent with the conditions of Equations (6.22) and (6.23).

The similarity based on Hamming distance introduced in Section 3.2 of Chapter 3 (repeated in Table 3.2) does not satisfy the condition of Equation (6.20), ***i.e.*** Assumption 2. In Minkowski family distances on the encoding of preference relations (see Equation (3.5) of Chapter 3), Assumption 2 is satisfied for L^ρ space while $\rho \ge 2$.

Table 6.2 – Hamming distance between preference relation types

	$\omega_{\succ}$	$\omega_{\prec}$	$\omega_{\approx}$	$\omega_{\sim}$
$\omega_{\succ}$	0	2	1	1
$\omega_{\prec}$	2	0	1	1
$\omega_{\approx}$	1	1	0	2
$\omega_{\sim}$	1	1	2	0

6.2.3 Application: Group decision making

As mentioned at the beginning of this chapter, the original objective of proposing WSD distance is to solve an issue of biased decision result confronting "indifference" in group decision making . For a better readability, we briefly review the process of group decision making with BFpref model.

For alternative pair $a_i, a_j \in A$, the decision making procedure consists mainly in two steps: BBA combination and decision, illustrated in Figure 6.3.

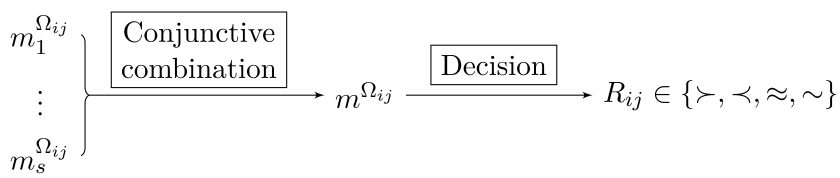


Figure 6.3 – Group decision making procedure

With absence of information interpreted as "incomparability" (in CS and KM models, see Section 3.2.2 of Chapter 3), the concensus decision strategies may converge to "absence of information". To address this issue, we apply conjunctive rule for the combination of BBAs. As mentioned in Section 5.3 of Chapter 5, conjunctive rule may cause high global conflict if the BBAs are simple on different singleton and number of sources is relatively large. The LNS-CR (Conjunctive Rule for Large Number of Sources) [ZMP18] is more adaptable for the aggregation of preferences without uncertainty because the BBAs are often separable, and this rule is equivalent to Strategy B in Section 5.3 of Chapter 5 when sources are simple BBAs.
In the decision step, we apply the minimum distance based strategy, repeated here:

$$\omega_{ij} = \underset{R \in \{\succ, \prec, \approx, \sim\}}{argmin} \left(d(m^{\Omega_{ij}}, \omega_R^0)\right)$$

An advantage of this decision rule is that it discriminates the weight of different singletons affected by WSD distance. This advantage is illustrated and discussed in Section 6.3.

6.3 Illustrative examples and experiments

In this section, we firstly illustrate the decision procedure by examples on synthetic data. Afterwards, in Subsection 6.3.2, we demonstrate the entire preference aggregation and decision procedure on imperfect data, applied on Sushi preference data.

6.3.1 On synthetic data

Different preference aggregation rules are compared on generated data. On alternatives a_i and a_j in quasi order with missing data, following specific preference cases are possible:

- $a_i \succ a_j$.
- $a_i \prec a_j$.
- $a_i \approx a_j$.
- $a_i \sim a_j$.
- preference information is not given, shorted as *vac* (vacuous).

With the number of each preference type variant, we compare the aggregated preference following different rules compatible with quasi orders:

- KM (JKM) model [KM01]
- CS model [RS93]
- Kamishima's model [Kam03a]
- BFpref model with Jousselme distance (BFJ)
- BFpref model with weighted singleton distance (BFwsd)

Experiment 1: Impact of missing data

As KM and CS model are not applicable on measuring incomparability relationship as mentioned in Section 3.2 of Chapter 3, thus, we generate preferences without incomparability but ignorance. All generated preferences are certain, thus their corresponding BBAs are categorical. For WSD distance, the similarity values between preference types are given in Table 6.3, with corresponding graphical representation in Figure 6.4.

The similarity setting in Table 6.3 takes extreme values, ***i.e.*** the equality conditions in Equation (6.23) are satisfied. In Table 6.4, the first column indicates the index of each experiment setting and columns $n_{\succ}$, $n_{\prec}$, $n_{\approx}$, $n_{\sim}$, and n_{vac} represent the number of agents giving the related preference relation.

Aggregation results are illustrated in Table 6.5. Experiment 1 shows several cases

Table 6.3 – Similarity values setting in Exp 1

	$\succ$	$\prec$	$\approx$	$\sim$
$\succ$	1	0	$\frac{1}{3}$	0
$\prec$	0	1	$\frac{1}{3}$	0
$\approx$	$\frac{1}{3}$	$\frac{1}{3}$	1	0
$\sim$	0	0	0	1

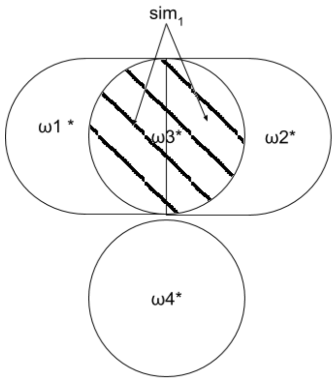


Figure 6.4 – Graphical illustration of similarity between singletons in Exp 1.

Table 6.4 – Experiment settings of Experiment 1

Exp. set	$n_{\succ}$	$n_{\prec}$	$n_{\approx}$	$n_{\sim}$	n_{vac}
1	5	4	4	0	7
2	5	3	3	0	20
3	5	5	4	0	7
4	5	5	5	0	0
5	5	5	5	0	20
6	5	5	0	0	0
7	5	0	5	0	0
8	5	0	5	0	20
9	5	0	4	0	20

where different results are obtained by different aggregation strategies. To simplify the expression, we use "MB" for "modified Borda rule", BFJ S, BFJ Ω respectively for BFpref model applying Jousselme distance for decision on singleton and on the whole discernment and BFwds S, BFwsd Ω respectively for BFpref model applying weighted singleton distance for decision on singletons and on the whole discernment. In column "combined BBA", non-zero values in the finally combined BBAs are illustrated (without value on $\emptyset$).

Following conclusions can be drawn:

1. In setting $N°1, 2, 3, 5, 8, 9$, KM or CS models return *vac*, indicating "missing data". Thus, distance based consensus rules are not capable to deal with preference data with high percentage of missing data. More specifically, KM model is

[1]Union sign $\cup$ indicates that the decision is on an imprecision set in 2^{Ω} while "or" indicates that multiple decisions are possible.

Table 6.5 – Results of Experiment 1[1]

Exp. set	KM	CS	MB	BFJ on S	BFJ Ω	BFwsd on S	BF wsd Ω	combined BBA
1	*vac*	$\approx$	$\succ$	$\succ$	Ω	$\approx$	$\succ \cup \prec \cup \approx$ or Ω	$m_1(\succ) = 0.184$ $m_1(\prec) = 0.131$ $m_1(\approx) = 0.131$ $m_1(\Omega) = 0.295$
2	*vac*	*vac*	$\succ$	$\succ$	Ω	$\succ$	$\succ \cup \prec \cup \approx$ or Ω	$m_2(\succ) = 0.184$ $m_2(\prec) = 0.131$ $m_2(\approx) = 0.131$ $m_2(\Omega) = 0.295$
3	*vac*	$\approx$	$\approx$	$\succ$ or $\prec$	Ω	$\approx$	$\succ \cup \prec \cup \approx$ or Ω	$m_3(\succ) = 0.164$ $m_3(\prec) = 0.164$ $m_3(\approx) = 0.118$ $m_3(\Omega) = 0.295$
4	$\approx$	$\approx$	$\approx$	$\succ$ or $\prec$ or $\approx$	Ω	$\approx$	$\succ \cup \prec \cup \approx$ or Ω	$m_4(\succ) = 0.148$ $m_4(\prec) = 0.148$ $m_4(\approx) = 0.148$ $m_4(\Omega) = 0.296$
5	*vac*	*vac*	$\approx$	$\succ$ or $\prec$ or $\approx$	Ω	$\approx$	$\succ \cup \prec \cup \approx$ or Ω	$m_5(\succ) = 0.148$ $m_5(\prec) = 0.148$ $m_5(\approx) = 0.148$ $m_5(\Omega) = 0.296$
6	$\succ$ or $\prec$	$\succ$ or $\prec$ or $\approx$	$\approx$	$\succ$ or $\prec$	$\approx$ or Ω	$\succ$ or $\prec$	$\succ \cup \prec$ or $\succ \cup \prec \cup \approx$ or $\succ \cup \prec \cup \sim$ or Ω	$m_6(\succ) = 0.25$ $m_6(\prec) = 0.25$ $m_6(\approx) = 0$ $m_6(\Omega) = 0.25$
7	$\succ$ or $\approx$	$\succ$ or $\approx$	$\succ$	$\succ$ or $\approx$	$\succ \cup \approx$ or Ω	$\succ$ or $\approx$	$\succ \cup \approx$	$m_7(\succ) = 0.25$ $m_7(\prec) = 0$ $m_7(\approx) = 0.25$ $m_7(\Omega) = 0.25$
8	*vac*	*vac*	$\succ$	$\succ$ or $\approx$	$\succ \cup \approx$ or Ω	$\succ$ or $\approx$	$\succ \cup \approx$	$m_8(\succ) = 0.25$ $m_8(\prec) = 0$ $m_8(\approx) = 0.25$ $m_8(\Omega) = 0.25$
9	*vac*	*vac*	$\succ$	$\succ$	$\succ \cup \approx$	$\succ$	$\succ \cup \approx$	$m_9(\succ) = 0.3$ $m_9(\prec) = 0$ $m_9(\approx) = 0.196$ $m_9(\Omega) = 0.247$

less robust than CS model confronting missing data while missing data has no influence on the results of Modified Borda, BFJ_S and BFwsd_S method.

2. In setting $N°1$ to 6, both BFJ_Ω and BFwsd_Ω risk to decide on Ω (vac) when the number of missing data is important. This jeopardizes the advantage of belief function based model mentioned above.

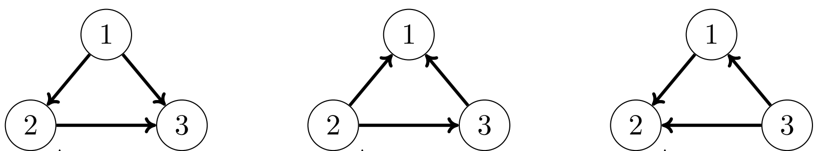
3. The modified Borda rule always returns a precise result. However, it is not able to distinguish $\approx$, $\sim$ and *vac*. Modified Borda is equivalent to merely count the number of $\succ$ and $\prec$ with plurality principle and the other type of preference are neglected.

4. Comparing BFpref model in setting 3 and 4, we could say that BFwsd_S gives a more pertinent result than BFJ_S. Given identical number of $\succ$ and $\prec$, it is more pertinent to be decided as $\approx$ than $\succ$ or $\prec$.

Experiment 2: Condorcet's paradox in preference aggregation

Condorcet's paradox resulting from belief function based preference model was discussed in [ZBM17], based on BFJ decision strategy. With the pertinent BFwsd strategy, Condorcet's paradox may be avoided in some cases.

Here is a simple example, three agents AGT_1, AGT_2, AGT_3 give the following uncertain preferences on three alternatives a_1, a_2, a_3 with the following values: correspondingly supporting the preference relations decided by both BFJ and BFwsd, illustrated in Figure 6.5.



Agent u_1 Agent u_2 Agent u_3

Figure 6.5 – Uncertain preferences of 3 agents

The aggregated preferences have the bba shown in Table 6.9. In this case, the decision rules BFJ and BFwsd return different final preference, as illustrated in Figure 6.6.

(a) BFJ rule (b) BFwsd rule

Figure 6.6 – Decision results from two BFpref consensus rules

Table 6.6 – Preference BBAs of u_1

	m($\omega_{\succ}$)	m($\omega_{\prec}$)	m($\omega_{\approx}$)	m($\omega_{\sim}$)
a_1, a_2	0.7	0	0.3	0
a_1, a_3	0.9	0	0.1	0
a_2, a_3	0.7	0	0.3	0

Table 6.7 – Preference BBAs of u_2

	m($\omega_{\succ}$)	m($\omega_{\prec}$)	m($\omega_{\approx}$)	m($\omega_{\sim}$)
a_1, a_2	0	0.9	0.1	0
a_1, a_3	0	0.7	0.3	0
a_2, a_3	0.7	0	0.3	0

Table 6.8 – Preference BBA values of u_3

	m($\omega_{\succ}$)	m($\omega_{\prec}$)	m($\omega_{\approx}$)	m($\omega_{\sim}$)
a_1, a_2	0.7	0	0.3	0
a_1, a_3	0	0.7	0.3	0
a_2, a_3	0	0.9	0.1	0

Table 6.9 – BBAs of aggregated preferences from 3 agents

	$m(\emptyset)$	m($\omega_{\succ}$)	m($\omega_{\prec}$)	m($\omega_{\approx}$)	m($\omega_{\sim}$)	$m(\Omega)$
a_1, a_2	0.147	0.186	0.075	0.166	0	0.426
a_1, a_3	0.147	0.075	0.186	0.166	0	0.426
a_2, a_3	0.147	0.186	0.075	0.166	0	0.426

On the aggregated BBAs in Table 6.9, the BFJ decision rule returns a “cycle” preference, which signifies a Condorcet’s paradox. However, BFwsd decision rule returns the “indifferent” relation among the three alternatives, which is a rational result.

A Condorcet’s paradox removing method based on BFpef model was proposed in [ZBM17]. The idea is based on replacing the “comparable” (i.e. “strict preference”, “inverse strict preference” or “indifference”) preference relation nearest to “incomparability” in the Condorcet’s cycle by an “incomparability” relation. This subject has been discussed in the previous chapter (Section 5.3.3 of Chapter 5), the details of the algorithm are described in Algorithm 2 and Algorithm 3 of Chapter 5.

The difference of results between BF_J and BF_WSD in removing Condorcet’s paradox procedure is caused by the difference in the values of distance function $d(m, \omega_{\sim}^{0})$.

6.3.2 Experiment 3: Conflicting preference aggregation on sushi preference dataset

In this experiment, we apply the BFpref model on Sushi dataset, which contains conflicting preference sources. We firstly introduce the data and detail the aggregation procedure. Then, we evaluate the aggregation quality by a specific "oiliness index", adapted to a common sense knowledge of Japanese sociology of food.

Conflict management on sushi preference dataset

Table 6.10 – Sushi preference dataset

	# agents	# alternatives	sparsity	conflict between rank and score	remark
East	3257	100	90%	3.45‰	more oily
West	1742	100	90%	3.44‰	less oily

In Sushi preference dataset, agents are categorized as east and west, indicating their geographical position. 3257 agents from the east region and 1742 from the west express their preferences over 10 sushis out of 100, in formats of both scoring and ranking. Thus, the sparsity of data is 90%. Owing to the fact that two sources preferences (ranking and scoring) are available, contradiction may exist. Detailed information on the dataset is given in Table 6.10. According to the Japanese sociology of food, eastern and western Japan has obvious differences. Generally speaking, the eastern Japanese habitats prefer more oily and more heavily seasoned food than the western Japanese habitats. The oiliness level of each type of sushi is also provided, denoted by $oil \in [0, 4]$, where 0 signifies the most oily taste and 4 the least (To make the data correspondent to a natural recognition, i.e. higher values implies more oily food, we transfer these values inversely, making 0 signify the least oily and 4 the most, as shown in Appendix D). We make group decision on different regions and verify if the result is coherent with this knowledge.
In sushi preference dataset, for an identical agent, the response in score and rank are cognitively dependent. Thus, as aforementioned in Section 5.3 of Chapter 5, the means rule (see Equation (2.25) in Chapter 2) is appropriate for preference aggregation on one identical agent. Extended from Figure 6.3 in Section 6.2.3, the procedure is given on Figure 6.7.

6.3.3 Evaluation

The evaluation on group decision making is tricky. Usually, the evaluation of the decision results are based on the effects in applications, such as feedback of service [AHVCH10], knowledge of ground truth [PC09], ***etc***. These criteria require additional information apart from the raw preference information. Other evaluation criteria measure the discrepancy or consensus degree between the ranking of final solution and the ranking

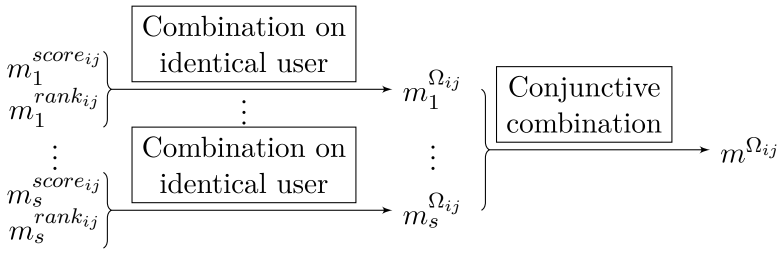


Figure 6.7 – Combination procedure for sushi preferences

solution of each expert, such as Deviation Index (DI), Violation index (VI), Best alternative coverage rate (BACR), introduced in [CLCH13]. We extend the criteria DI to be compatible with indifference relation. Average deviation index ADI measures the consensus level of each agent s and the aggregated preference by average of Fagin distance $\tau(\cdot)$ [FKS03a], with distance between indifference and strict preference valued as 0.5. The absent preferences from an individual agent are not taken into account. Formally:

$$ADI = \frac{1}{|S|} \sum_{s \in S} \tau(\sigma_{agg}, \sigma_s) \tag{6.25}$$

where σ_{agg} denotes the aggregated preference order, σ_s the preference order of agent s. The average deviation index of four rules on two regions is given in Table 6.11. In

Table 6.11 – Aggregated conflicting Sushi preferences with ADI

	ADI order BF J	ADI score BF J	ADI order BF WSD	ADI score BF WSD
East Japan	0.2685	0.2395	0.1462	0.1172
West Japan	0.2661	0.2417	0.1346	0.1103

Table 6.11, we can clearly conclude that decision rule of BF_WSD gives a more consensus aggregation result than BF_J, both on preferences in format of score and in format of rank. It should be pointed that the ADI is not a convincing homogeneous evaluation criterion. If the preference aggregation method is based on the minimization of a preference order distance, denoted as d_{agg}, the best result in the evaluation always corresponds to d_{agg} applied in the aggregation.

We also evaluate the preference aggregation results with the ground truth. For sushi preference dataset, having the knowledge that Japanese from east region usually prefer more oily sushi, we reckon that if the aggregated preferences is able to discriminate the oiliness largely from different region, the aggregation rule is appropriate. We define an average oiliness index (AOI) on the whole aggregated preference.

Given the oiliness level of sushi a_i, denoted as $oil(a_i)$, oiliness index AOI on a set of sushi A_{sushi} is simply calculated by

$$AOI(A_{sushi}) = \frac{1}{|A_{sushi}|} \sum_{a_i \in A_{sushi}} oil(a_i) \tag{6.26}$$

We compared AOI on top-k sushis respectively from east and west Japan. The comparison is calculated by the difference:

$$AOI_{diff} = AOI(A^{East}_{sushi}) - AOI(A^{West}_{sushi}) \tag{6.27}$$

where $AOI(A^{East}_{sushi})$ and $AOI(A^{West}_{sushi})$ respectively represent AOI on aggregated sushi orders in East and West Japan.

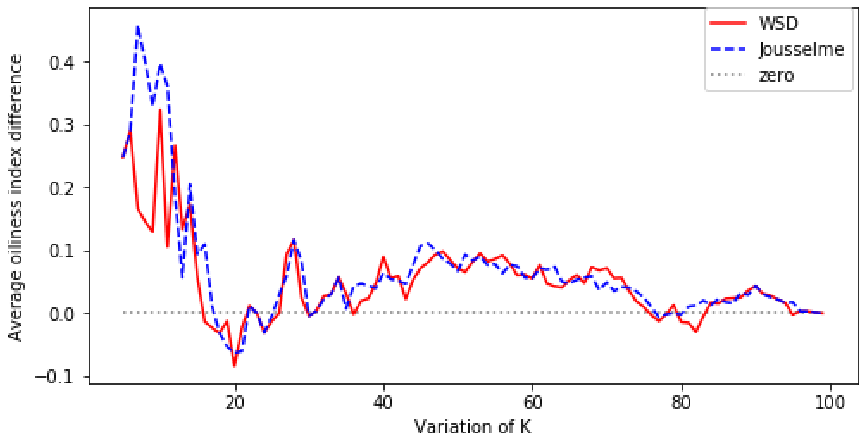


Figure 6.8 – Sushi AOI difference between East Japan and West Japan

In Figure 6.8, k (in top-k) varies from 1 to 100. Axis 0 represents the case that habits from East and West Japan share equal tastes in terms of oiliness. From the definition in Equation (6.27), a positive value of AOI_{diff} indicates that top-k in east Japan is more oily than the west Japan and vice versa for a negative value. It can be figured AOI_{diff} is positive globally in all values of k both in decision rule based on Jousselme distance and WSD distance, especially in top-10 sushi. This result supports the knowledge that people in east Japan prefer more oily sushis. However, AOI_{diff} is negative around $k = 20$. This is caused by the sushi ranked the 19th and 20th in east Japan (sushi Hamachi and tekka maki) are less oily than those in west Japan (sushi Uni and Botanebi) (see Appendix D and E). However, the knowledge that people in east Japan eat more oily is not strictly defined and the fact that people in west Japan prefer several oily sushi is tolerable.

The entire ranking list in east and west Japan as well as additional information on these sushi are provided in Appendix D and E.

6.4 Conclusion

It is well accepted that preference relations consist of three types of binary relations: “strict preference”, “indifference” and “incomparability”. However, disagreements exist in the interpretation of “incomparability”. Some works interpret it as “not compared” resulted from the absence of information and other work as “not able to compare” because of some causes such as conflicts. Supported by the original definition, we analyzed different interpretations and accepted the latter one. We demonstrated that the “incomparability” is exclusive with other relations and the case of “not compared” caused by the absence of information is actually the union of all possible relations, ***i.e.*** total ignorance.

BFpref model for preferences with imprecision and ignorance based on the theory of belief functions is able to express the different interpretations in one BBA. The unobserved preferences are considered as total ignorance, and the relation of “incomparability” is considered as an exclusive relation to “strict preference” and “indifference”. This model respects strictly their original definitions.

In decision making, given a BBA, minimum of distance to categorical BBA and maximum on pignistic are widely accepted. We proved that the decision making strategies of maximum pignistic and minimum on Jousselme distance on singletons are equivalent. However, existing metrics for BBA similarity measure take the singletons equally, ***i.e.*** the dissimilarity are not weighted. For preferences, it is more relevant to consider that the distance between “strict preference” and “inverse strict preference” is larger than that between “strict preference” and “indifference”.

To solve this issue, we proposed a novel similarity measure metric (named Weighted Singleton Distance, WSD) for BBAs, taking the similarity between singletons into account. WSD is de facto an extended version of Jousselme distance, it returns to Jousselme distance when similarity between different singletons are equal. WSD is more relevant than Jousselme distance for BFpref model because it is able to differentiate the similarity between different types of preference relations. We applied WSD metric for decision making strategy, especially for Condorcet’s paradox removal. We compared strategies based on WSD and Jousselme distance and concluded that WSD is more pertinent.

In the experiment part, the similarity values between different preference relations are assigned by extreme values, i.e. the equality case in Equation eq:valuecondition. Indeed, the similarities between four preference relations are not definitively defined. As introduced in Section 3.2 of Chapter 3, various similarity values may be adaptable, depending on the application cases. In this thesis, only the limitation conditions. The selection of the similarity value can be affected by additional knowledge upon the using cases, or by learning methods.

Besides, in Chapters 3 and 4, we introduced that distance is applicable in both preference aggregation and preference learning in termes of preference manangement. In the scope of this work, we applied WSD distance in preference aggregation. However, its applications in classification are not justified. In the cases of certain preferences, WSD is equivalent to Kendall’s τ distance in case of total order and to Fagin’s distance in case of weak order. Thus, it is pertinent to apply WSD for classification on complete preferences both certain [FKS03b] or uncertain [ZBM18a] cases. Nevertheless, for incomplete preferences, where the impact of ignorance is important, it is reasonable to doubt the appropriateness of WSD. In the future work, we will study on the classification applications based on similarity between imperfect preferences.

In the next chapter, we study the possibility of clustering on evidential preference learning with BFpref model.

Chapter 7

Clustering on evidential preferences

In previous chapters, we have studied the aggregation issues on BFpref model for evidential preferences. As mentioned in Chapter 4, preference learning is a promising and popular domain. It can be applied in various applications such as community detection in social network, or preference elicitation in recommendation systems. In this chapter, we introduce our contribution on the application of BFpref model in preference learning.

Based on BFpref model, a distance for complete preference structures with uncertainty was proposed in Chapter 6, solving an issue of decision on singletons with distinctive weights. However, the distance in learning methods required more properties. For example, the measure of distance between two pieces of total ignorance are regarded as identical in distance measure, but to classify such two pieces of knowledge into one group is abitrary. In Section 7.1, we firstly introduce the issues learning on evidential objects and propose properties required. Based on these properties, we found a paradox in learning on evidential objects with distances. Afterwards, in Section 7.2, we applied EK-NNclus method [DKS15b] for clustering on data of complete evidential preferences confronting multiple sources. The clustering results on BFpref models are compared with a simple mean value strategy in Section 7.4.

When applying EK-NNclus method, a method for the determination of the number of clusters k was proposed, based on silhouette coefficient score [Rou87] and elbow method [Tho53]. However, this work is not yet applicable on clustering of evidential preferences, since it require the calculation of centroïd. The experiment and a demonstrative guide of this k determination method is given in Section 7.5, after the experiment part of evidential preference clustering.

7.1 Scientific issues on distances for evidential preferences

The clustering process is to segment objects into multiple groups (called clusters) based on their similarity such that objects in the same cluster share more degree of similarity to each other than to those in other clusters. Usually, this similarity is measured by distance. A most simple method is to centroïd based clustering such as k-means [M^+67].

7.1.1 Distance in k-means

In the first phase, we tend to cluster the $\mathcal{AGT}$ represented by order set $\mathcal{OD}$ with a k-means based algorithm.
Generally, a k-means based algorithm can be regarded as three principal steps:

1. Initialization: randomly choose k initial centroïds
2. Assignment step: Compute cluster labels and distance to the corresponding cluster center of every object
3. Update step: Calculate the center of clusters

For M orders on N alternatives, the objects are represented by a 3-D tensor of dimension $M \times V \times 2^{\Omega}$, where

$$V = \frac{N(N-1)}{2} \tag{7.1}$$

denotes the number of possible alternative pairs, 2^{Ω} the size of mass vector with discernment frame of Ω.

7.1.2 Consistence between distance and combination rules

For evidential objects represented by BBAs, combination rules on BBAs are applied in the update step while conflicts measure methods are used in assignment step (calculating of the inertia). Given a set of evidential elements (represented by BBAs) on discernment framework Ω

$$\mathcal{E}(\Omega) \subseteq [0,1]^{2^{\Omega}}, \tag{7.2}$$

a similarity measure d is defined as:

$$d : \mathcal{E}(\Omega) \times \mathcal{E}(\Omega) \to \mathbb{R}_{\geq 0} \tag{7.3}$$

(see Definition 3.1.5 in Chapter 3). Concerning our case, following conditions are supposed to be satisfied:

1. **Metric consistency:** For the objects that a k-means algorithm is theoretically correct, i.e. the iteration result converges, the combination rule $\bigodot$ and the similarity measure d must be consistent. Formally, the combined BBA m_{comb} is given by:

$$m_{comb} = \bigodot_{i=1}^{k} m_i \tag{7.4}$$

such that $\forall m' \in \mathcal{E}(\Omega), m' \neq m_{comb}$,

$$\sum_{i=1}^{k} d(m_k, m') \geqslant \sum_{i=1}^{k} d(m_k, m_{comb}) \tag{7.5}$$

A stronger property for similarity measure is metric (see Definition 3.1.1 in Section 3.1), with properties repeated here: $\forall m_x, m_y, m_z \in \mathcal{E}(\Omega)$

- Non-negativity: $d(m_x, m_y) \geq 0$
- Identity of indiscernibles: $d(m_x, m_y) = 0 \Leftrightarrow m_x = m_y$
- Symmetry: $d(m_x, m_y) = d(m_y, m_x)$
- Triangle inequality: $d(m_x, m_z) \leq d(m_x, m_y) + d(m_y, m_z)$

2. **idempotence:** The combination rule $\bigodot$ must be idempotent, i.e.

$$\forall m \in \mathcal{E}(\Omega), m \bigodot m = m. \tag{7.6}$$

3. **Surjectivity of combination result:** The combination rule is surjective. i.e. For a given set of BBAs, there exist one and only one combination result. Hence, the equality in Inequation (7.5) is not possible.

4. **Ignorance neutrality:** The knowledge representing ignorance is supposed to play a "neutral role" in the combined result. Formally, for combination rules $\bigodot$.

$$\forall m \in \mathcal{E}(\Omega), m \bigodot m_\Omega = m. \tag{7.7}$$

Indeed, the property of "idempotence" is a corollary of "metric consistence". The proof is given in Appendix B.

7.1.3 Incompleteness of preference orders

Another difficulty exists in the dissimilarity measure between preferences in quasi-orders, where "incomparability" may exist, interpreted as "missing data" (or "incompleteness") in this context. However, none of the metric in Chapter 3 is able to measure the dissimilarity between preferences with such "incompleteness". An extreme case is given: $\mathcal{A}_1, \mathcal{A}_2$ are two subsets of comparable alternative set such that $\mathcal{A}$: $\mathcal{A}_1 \subset \mathcal{A}, \mathcal{A}_2 \subset \mathcal{A}$, and $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$.

Two preferences σ_1, σ_2 in a quasi-order are respectively on $\mathcal{A}_1$ and $\mathcal{A}_2$. In this case, we need to measure the dissimilarity between two preferences with no alternatives in common, which may exist in quasi-orders.

7.1.4 Some combination rules and their properties

We have studied several combination rules and similarity measure methods. However, none of them works well for the following reasons.

- Dempster's combination rule. This rule is conjunctive, but not idempotent
- Cautious rule. This rule is idempotent, but the ignorance is not neutral.

- Minimization of Jousselme's distance. This rule is idempotent but not conjunctive. With most of the objects represented by m_Ω, the combination result converge to m_Ω.
- Klein's idempotent combination rule [KDC18]. This rule is both idempotent and conjunctive. However, it is still impossible to define a similarity measure consistent with this rule, generally proved by Proposition 2 afterwards.

Indeed, concerning the properties in Section 7.1, a impossibility theorem proposed below.

Proposition 2. On set of evidential elements $\mathcal{E}(\Omega)$ (represented by BBAs), given a similarity measure $d : \mathcal{E}(\Omega) \times \mathcal{E}(\Omega) \to \mathbb{R}_{\geq 0}$, there is no such d that meets the properties of ***metric consistence***, ***uniqueness of combination result***, and ***ignorance neutrality***.

The demonstration is given in Appendix B.

Without solving this issue on incomplete preferences, we started with the complete preferences, ***i.e.*** the partial orders.

7.2 A proposed method for clustering agents according to evidential preferences

In this section, we explain how the agents are represented and clustered from two sources of preferences. The clustering procedure is straightforward, it is concisely illustrated in figure 7.1. The first block concerns the representation of agents and modeling of

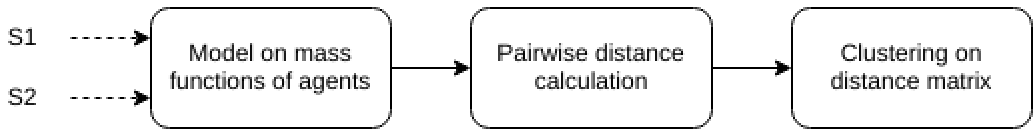


Figure 7.1 – Flowchart of preference clustering

mass functions from two preference sources S_1 and S_2. The second block concerns the calculation on the similarity between agents. The third block concerns the clustering algorithm, we used EK-NNclus algorithm in our work, explained later.

7.2.1 Representation of agents

We consider the case that a group of K agents expressing their preferences between each pair of alternatives from the set $\mathcal{A}$ of size N. Therefore, the preference of an agent agt_k, denoted by σ_k, is represented by a mass function on all possible alternative pairs:

$$\sigma_u := \Big[m_{1,2}; \quad m_{1,3}; \quad \ldots; \quad m_{1,N}; \quad m_{2,3}; \quad \ldots; \quad m_{N-1,N} \Big] \tag{7.8}$$

Hence, for N alternatives, the representation of an agent is made up by $\frac{N(N-1)}{2}$ mass functions.

7.2.2 Modeling of mass functions

In our model, we take advantage of the possibility of expressing on ignorance in the framework of BFpref (see Equation (5.1) in Section 5.2). Given two preference order sources $\sigma_{n,1}$, $\sigma_{n,2}$ from one agent agt_n, we interpret the Fagin distance $d_F(\sigma_1, \sigma_2)$ as ignorance degree when conflict encountered. That is to say: for a_i, a_j $(i < j)$ in both $\sigma_{n,1}$, $\sigma_{n,2}$ of agent agt_n, the mass function value is given according to following conditions:

1. **Case 1**: a_i, a_j are in the same relation in both σ_1, σ_2 (say $a_i \succ a_j$), m_{ij} is a categorical mass function on the corresponding element ($m_{ij}(\omega_\succ) = 1$).
2. **Case 2**: a_i, a_j are in the conflicting relations in σ_1 and σ_2, respectively denoted as ω_{ij,σ_1}, $\omega_{ij,\sigma_2} \in \Omega_{ij}$, $\omega_{ij,\sigma_1} \neq \omega_{ij,\sigma_2}$ (say $a_i \succ a_j$ in σ_1 while $a_i \approx a_j$ in σ_2, thus, $\omega_{ij,\sigma_1} = \omega_\succ, \omega_{ij,\sigma_2} = \omega_\approx$), the mass function values are given by:

$$\begin{aligned} &m_{ij}(\Omega) = d_F(\sigma_1, \sigma_2) \\ &m_{ij}(\omega_{ij,\sigma_1}) = m_{ij}(\omega_{ij,\sigma_2}) = (1 - d_{Fagin}(\sigma_1, \sigma_2))/2 \end{aligned} \tag{7.9}$$

   where d_{Fagin} denotes the Fagin's distance (see Equation (3.14) in Chapter 3) between two orders, with penalty value $p = 0.5$.

7.2.3 Similarity between different agents

The dissimilarity measure is based on Jousselme distance [JGEB01] for mass functions. Given two mass functions modeling preference relations between alternatives a_i and a_j from agents agt_1, agt_2 expressing preference orders σ_1, σ_2, we denote Jousselme distance as $d_J(m_{ij,\sigma_1}, m_{ij,\sigma_2})$. The dissimilarity between two agents' preferences is denoted via Jousselme distance as:

$$d(\sigma_1, \sigma_2) = \sum_{j=1}^{k} \sum_{i=1, i<j}^{k} d_J(m_{ij,\sigma_1}, m_{ij,\sigma_2}) \tag{7.10}$$

Where $m_{ij,\sigma}$ denotes the mass function of alternative pair (a_i, a_j) according to the order σ. Therefore, a normalized distance is given by

$$d_{Normalize}(\sigma_1, \sigma_2) = \frac{1}{Nb_{total}} d(\sigma_1, \sigma_2) \tag{7.11}$$

where $Nb_{total} = \frac{N(N-1)}{2}$, is the amount of all alternative pairs.

Indeed, normalized distance degrade to Kendall distance when preferences are certain, ***i.e.*** all BBAs are categorical. A proposition is therefore concluded:

Proposition 3. The normalized distance function defined in Equation (7.11) is equivalent to Kendall distance when the preference orders are total and crisp.

The demonstration is given in Appendix C. The clustering process is therefore applied on this normalized distance function. In the next section, we focus on the selection of clustering algorithms.

7.3 Introduction to an unsupervised classifier–EK-NNclus

For similarity spaces in which only pairwise distances are given (such as Kendall distance), the centroïd of several agents is a metric k-center problem and is proved to be NP-hard. Moreover, uniqueness of the centroïd is guaranteed only when the similarity space is a Riemannian manifold [FVJ09]. Therefore, we avoid using clustering methods requiring the calculation of centroïd, such as k-means.
Various clsutering algorithms adaptable for pairwise distances exist, such as connectivity based model and density models. The comparison of different clustering methods is not in the scope of this thesis, we applied EK-NNclus method [DKS15b] as clustering approach. A brief introduction is given below.

7.3.1 EK-NNclus algorithm

EK-NNclus [DKS15b] is a clustering algorithm based on the evidential k-nearest-neighbor classifier (Ek-NN classifier) [Den95]. Pairwise distances is sufficient for k-nearest-neighbor searching. Thus, EK-NNclus is independent to the calculation of centroïd. EK-NNclus starts from an initial random set of clusters, and iteratively reassigns objects to clusters using Ek-NN classifier. The algorithm converges to a stable status of partition. For each object, its membership to clusters is described by a mass function in a framework of each cluster and the whole set of clusters (***i.e.*** ignorance). For a set of objects

$$\mathcal{X} = \{x_1, \ldots, x_K\},$$

given a matrix of pairwise distances $D = (d_{ij})$, where d_{ij} denotes the distance between objects x_i and x_j. In EK-NN classification rule, a discernment frame is defined for each objects belonging to clusters $\mathcal{C} = \{c_1, ...c_C\}$ as:

$$\Omega^{Ec} = \{\omega_1, \ldots, \omega_C\} \tag{7.12}$$

The BBA of degrees on each cluster of object x_i is denoted as m_i^{Ec}. The set of k nearest neighbors for object x_i is denoted as $NK(i)$.

According to [DKS15b], the procedure of EK-NNclus can be briefly divided into the following parts:

- **Preparation** Calculate the mass value η_{ij} of event that x_j is in the k-nearest neighbors of x_i based on d_{ij} by a non-increasing mapping function $\phi(d_{ij})$, defined as:

$$\eta_{ij} = \begin{cases} \phi(d_{ij}), & \text{if } j \in NK(i) \\ 0, & \text{otherwise} \end{cases}$$

- **Initialization** Initialize the labels of each object randomly. The authors of [DKS15b] suggest that number of clusters c can be set to the number of objects n if n is not too large. A binary membership function $inClus_{i,c}$ on object x_i and cluster c is also initialized randomly, where $inClus_{i,c} = 1$ implies x_i belongs to cluster c and 0 otherwise.

- **Iteration** Randomly reorder all objects. Then, for every object x'_i in the new order, calculate the plausibility of belonging to each cluster. Plausibility of x_i belong to cluster c is calculated by:

$$u_{i,c} = \sum_{j \in NK(i)} v_{ij} \times inClus_{i,c}, c \in \mathcal{C} \tag{7.13}$$

  where v_{ij} is calculated from η_{ij} by:

$$v_{ij} = -ln(1 - \eta_{ij}). \tag{7.14}$$

  Then, assign x'_i to the cluster with the highest plausibility.

$$inClus_{i,c} = \begin{cases} 1, & \text{if } u_{i,c} \text{ is the maximum in } \mathcal{C}, \\ 0, & \text{otherwise} \end{cases}$$

- **Convergence condition** The iterations stop when the labels of all objects are stable.

After the convergence condition is reached, the final BBA of clustering result on x_i is calculated by:

$$m_i^{Ec} = \bigcirc\!\!\!\!\cap_{j \in NK(i)} m_{ij}^{Ec} \tag{7.15}$$

where

$$m_{ij}^{Ec}(\omega_k(j)) = \eta_{ij} \tag{7.16}$$

$$m_{ij}^{Ec}(\Omega^{Ec}) = 1 - \eta_{ij} \tag{7.17}$$

and $\bigcirc\!\!\!\!\cap$ the conjunctive combination rule.

In this procedure, the number of k at the preparation step has a vital impact on the clustering results. If k is too small, the matrix of η becomes sparse. In this case, the iteration times are very few and the clustering result depends highly on the initialization step, which is usually random. If k is too large, two objects far away from each other may be considered as in the same neighborhood. This may have two consequences:

1. The computation time becomes important.

2. Objects naturally in different clusters may be targeted as in the same one, causing a sub-estimation of number of clusters.

Therefore, the determination of k is important to guarantee a good quality of clustering.

The optimal k varies with the scale of the data. There is no identical k for datasets even respecting identical distribution. Therefore, the determination of k is necessary for every clustering analysis problem. The determination of k is two-fold. An optimal k in EK-NNclus should:

1. Cluster the data into an optimal number of clusters.
2. Return a result with high quality, close to the ground truth density or knowledge if they are known.

There are already some often-applied methods to determine the number of clusters C, usually these methods are related with evaluation criteria, such as silhouette coefficient [Rou87]) optimization, elbow method [Tho53] and information criterion approach [GHLR01]. By combining Adjusted Rand Index (ARI), Elbow method and silhouette score, we proposed a k determination strategy for EK-NNclus method in [ZBM18b]. We have to admit that this method is not applied for clustering on evidential preferences because it still needs the calculation of centroïd. Therefore, a brief introduction on this work is given below with experiment part in Section 7.3.3.

7.3.2 Evaluation criteria for clustering

In an ideal case where ground truth knowledge is given, it is natural that the clustering results more consistent to the ground truth correspond to better methods. This consistence between two clustering result is usually evaluated by Adjusted Rand Index (ARI) [HA85]:

ARI

Given the knowledge of the ground truth class assignments L_t and our clustering algorithm assignments of the same samples L, Adjusted Rand Index measures the similarity of the two assignments, with chance normalization, ignoring permutations. The ARI is calculated as follows:
Let us define α and β as:

- α: the number of pairs of elements that are in the same set in L_t and in the same set in L
- β: the number of pairs of elements that are in different sets in L_t and in different sets in L

The raw (unadjusted) Rand index is then given by:

$$RI = \frac{\alpha + \beta}{C_2^{|\mathcal{X}|}} \tag{7.18}$$

where $C_2^{|\mathcal{X}|}$ is the total number of possible pairs in the dataset (without ordering). However the RI score does not guarantee that random label assignments will get a value close to zero (especially if the number of clusters is in the same order of magnitude as the number of samples). To counter this effect we can discount the expected RI ($E[RI]$) of random labels by defining the adjusted Rand index as follows:

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]} \tag{7.19}$$

Obviously, the range of ARI is $\lfloor 0, 1 \rceil$ while 1 relates to the best case and 0 the worst.

Silhouette coefficient

If the ground truth labels are not known, evaluation must be performed using the model itself. The silhouette coefficient is such an evaluation criterion. A higher silhouette coefficient score relates to a model with better defined clusters. The silhouette coefficient is defined for each sample and is composed of two scores:

- $\bar{d}_{intra}$: the mean distance between a sample and all other points in the same class (intra-class).
- $\bar{d}_{inter}$: the mean distance between a sample and all other points in the next nearest class (inter-class).

The silhouette coefficient sil for a single sample is given as:

$$sil = \frac{\bar{d}_{inter} - \bar{d}_{intra}}{max(\bar{d}_{inter}, \bar{d}_{intra})} \tag{7.20}$$

The silhouette coefficient for a set of samples is given as the mean of the silhouette coefficient for each sample. From the definition above, for a dataset, silhouette coefficient close to 1 indicates a satisfying clustering result while 0 a bad one.

Elbow method

The elbow method [Tho53] applies the distortion as a criterion for clustering result. The rule is simple: among a set of different numbers of clusters C, one should choose a number $c \in \mathcal{C}$, such that $c + 1$ clusters do not give a much better modeling of the data. Given object set $\mathcal{X}$ in c clusters, we denote the center of clusters by $\mu_1, \mu_2, \ldots, \mu_c$. The quality of the modeling is measured by the distortion J of the clustering, calculated by:

$$J(c, \mu) = \frac{1}{|X|} \sum_{x_i \in \mathcal{X}} (\min_{j=1}^{c} (x_i - \mu_j)^2) \tag{7.21}$$

where $|X|$ denotes the size of objects set. Therefore, c can be subjectively determined with the help of a distortion plot helps, illustrated in the experiment part of Section 7.5.2.

A disadvantage of elbow method is that the “elbow" cannot always be unambiguously identified [KJS96]. The observation of the “elbow" is subjective because “a cluster that does not give a much better modeling of the data" cannot be justified quantitatively. Another inconvenience of the elbow method is that the calculation of distortion is based on the centroïd of each cluster. This jeopardizes the property that EK-NNclus is independent of the calculation of centroïd.

7.3.3 A k determination strategy

The idea of k determination is simple: an optimal k in EK-NNclus should return a high quality clustering result. Given a dataset, the quality of clustering can be easily evaluated if knowledge of ground truth is provided. A high value of ARI between clustering result and the ground truth implies a good clustering quality. However, in most cases, the ground truth is absent. The results of clustering are often evaluated by how well different clusters are separated. Silhouette coefficient is such a criteria and it is often strongly correlated with ARI. The correlation is plotted in the Section 7.5.1. However, to determine k only by silhouette coefficient is still risky. Fewer clusters may sometimes return a higher silhouette coefficient (example illustrated in Section 7.5.1 and Figure 7.12b). Thus, other conditions are needed. Elbow method is used as the second criterion to avoid that too few clusters are detected. The strategy is straightforward. From the intersection of the set of k ($\mathcal{K}_c$) corresponding to the best c and the set of k ($\mathcal{K}_{sil}$) corresponding to relatively high silhouette coefficient, the interval of values of k is obtained. We denote a set of all possible k by $\mathcal{K}$. A proper subset of k is therefore refined by: $\mathcal{K}_{refine} = \mathcal{K}_c \cap \mathcal{K}_{sil}$. We define a silhouette efficient function $f_{sc}(k)$, implying the silhouette coefficient of the clustering result with k in EK-NNclus algorithm. Thus, the optimal k is given by:

$$k = \underset{k \in \mathcal{K}_{refine}}{\arg\max}\,(f_{sc}(k)). \tag{7.22}$$

Because of the facts that the elbow method is a subjective method and that "relatively high silhouette coefficients" are also subjectively defined, both $\mathcal{K}_c$ and $\mathcal{K}_{sil}$ are indefinite sets. Thus, if $\mathcal{K}_{refine} = \emptyset$, we can extend $\mathcal{K}_c$ by softer condition or $\mathcal{K}_{sil}$ by lower threshold to obtain a non empty $\mathcal{K}_{refine}$.

As the determination of k in EK-NNclus method does not relate to the main subject of this thesis, we put demonstrative tutorial examples as well as experiments in Section 7.5.

Discussions on k determination method

Although combining silhouette coefficient and elbow is helpful in k determination, it is not applicable in our case, i.e. clustering on evidential preference. This is due to some short-comings conducted by elbow method. Firstly, the distortion requires the calculation of centroïds of clusters, which neutralizes an advantage of EK-NNclus: EK-NNclus is adaptable to pairwise distances and independent to the calculation of centroïd. Besides, the determination of c by elbow method is subjective and can be sometimes ambiguous. In the future, we can replace elbow method by centroïd-independent c determination method, making the strategy more adaptable.

In our work of clustering on evidential preferences, the determination of k in EK-NNclus is merely based on maximization of silhouette coefficient.

7.4 Experiments of clustering on evidential preferences

Although BFpref model was originally designed for preferences under uncertainty, we are still curious about its quality for clustering on certain preferences. Thus the clustering quality of our model can be divided into two aspects: on crisp preferences and on uncertain preferences.

We tested different metrics on synthetic certain and uncertain preferences. We also compared different metrics on a real world certain preferences from SUSHI dataset [Kam03b].

In the following parts, we introduce the method of generating synthetic preferences and compare the clustering quality of different metrics. To simplify the experiments, all preferences are expressed in a space of 10 alternatives.

7.4.1 Crisp preferences

The experiments in this section are executed on crisp preferences, ***i.e.*** preferences without uncertainty or imprecision, from both synthetic data and real world data.

Synthetic data

Certain preferences are those who are from non-conflicting sources. In this case, we only consider and generate one source of preferences. To study the clustering quality, we firstly generate preferences with different ranges to their centroïds.

The data is generated by Algorithm 4. By increasing the number of switching operations T, we obtain clusters with different density. To avoid random errors, we generate distinctive preference sets 10 times and take the average value of ARI and silhouette score. Besides, the optimal parameter k in EK-NNclus algorithm varies with the size of data and distribution of the samples. We test on various k and choose the one that returns the largest ARI and average silhouette coefficient[1] as our result.

In figures 7.2, 7.3, 7.4, ARI and silhouette coefficient performed on generated data with neighbors in various ranges (switch time from 1 to 3) and different sizes (neighbor size[2] varies from 10 to 100) are illustrated.

According to these results, one can conclude that the BFpref model and Kendall distance have equivalent good quality both in terms of ARI and silhouette score, while Euclidean distance always has a poor quality. A high value in ARI usually corresponds to a high silhouette score, signifying a good clustering result.

Real data

SUSHI preference dataset [Kam03b] is collected from a survey on Japanese consumer preferences over different sushis. It has a data set containing 5000 complete strict rank orders (***i.e.*** total orders) of 10 different kinds of sushi. We applied these three metrics in clustering on real data of Sushi Preference Data Set. Figure 7.5 illustrates

[1] Without special remark, we use term “silhouette coefficient” for “average“ value on set of samples by default.

[2] By saying neighbor size, we mean the number of samples in each cluster.

Algorithm 4 Generate preferences in $|C|$ clusters

Input: Cluster number C
Switch time T
neighbour size NS
Alternative size in each order N
Output: C clusters of preferences

```
// Centroids initialization
randomly generate centroïd c_1 of N elements.
for i_c in 2 : C do
  dist_max = 0
  for s in 1 : 5000 do
    randomly generate preference o_s of N elements
    dist_sum = \sum_{i=1}^{i_c-1} d_{Kendall}(o_s, c_i)
    if dist_sum > dist_max then
      dist_max = \sum_{i=1}^{i_c-1} d_{Kendall}(o_s, c_i)
      centroïd c_{c_i} = o_s
    end if
  end for
end for
Generate neighbors
for each centroïd o_c do
  for ns in 1 : NS do
    for t in 1 : T do
      randomly generate index i, j
      exchange ranking order of a_i, a_j in o_c, making a new order
    end for
  end for
end for
```

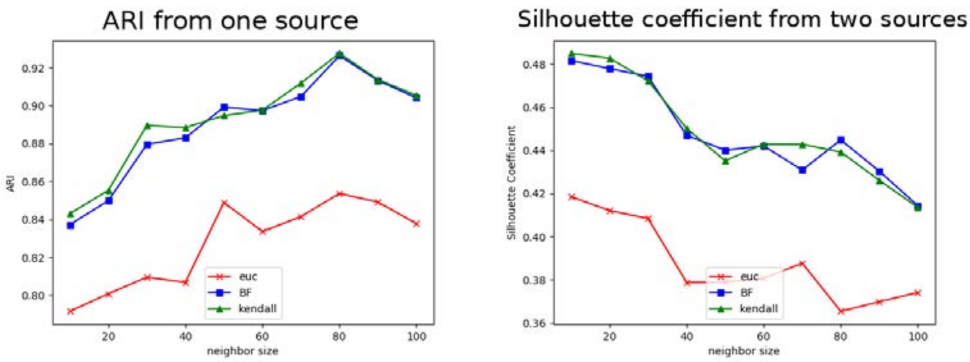


Figure 7.2 – ARI and silhouette coefficient, switch = 1

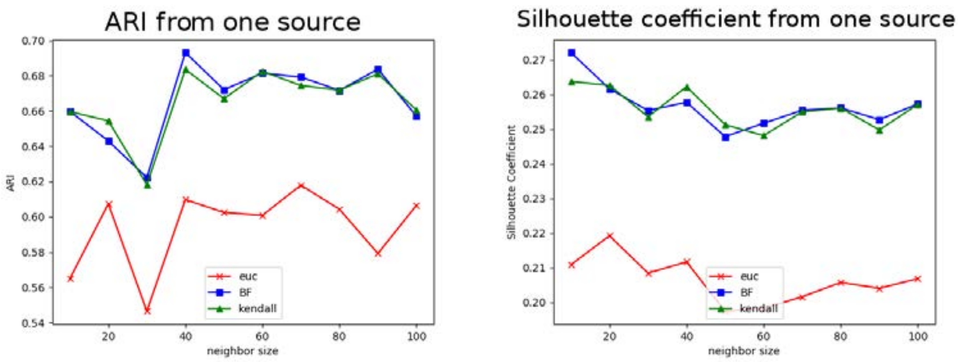


Figure 7.3 – ARI and silhouette coefficient, switch = 2

silhouette plots of clusters with different metrics[3]. Kendall distance and BFpref model have similar quality. Euclidean distance has a relatively poor quality. This result is consistent with the synthetic data in Figures 7.2, 7.3 and 7.4.

Among the three metrics, none of them has an absolutely high silhouette score (larger than 0.5). This is due to the quality of the data. SUSHI dataset doesn't guarantee the existence of obvious communities among the agents.

7.4.2 Uncertain preferences

In this part, we suppose a case that two preferences are given with different representations: ranking and score. The ranking preferences are generated in the same way as in subsection 7.4.1. Scores are generated by the following steps: scores range from 1 to 5 are generated respecting a given rank preference. In this way, indifference relations are introduced, causing conflicts between two preference sources. Given a ranking preference σ_r of 10 alternatives a_1 to a_{10}, the scores are generated by the following rules:

[3]As different K in EK-NNclus algorithm returns different clustering results, we compare clustering result who returns relatively high silhouette coefficients.

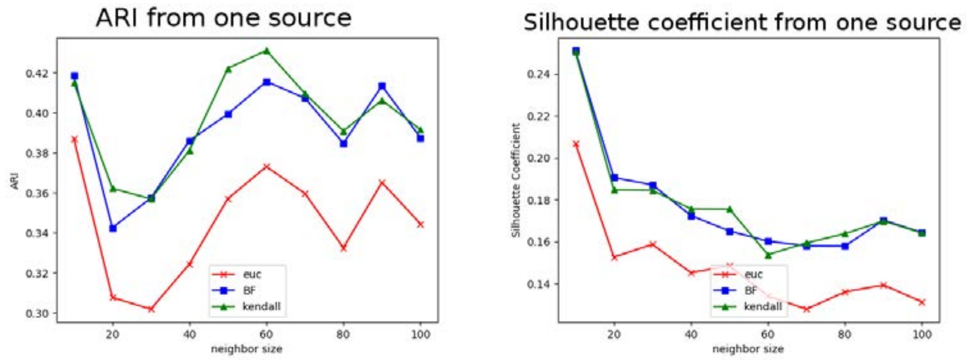


Figure 7.4 – ARI and silhouette coefficient, switch = 3

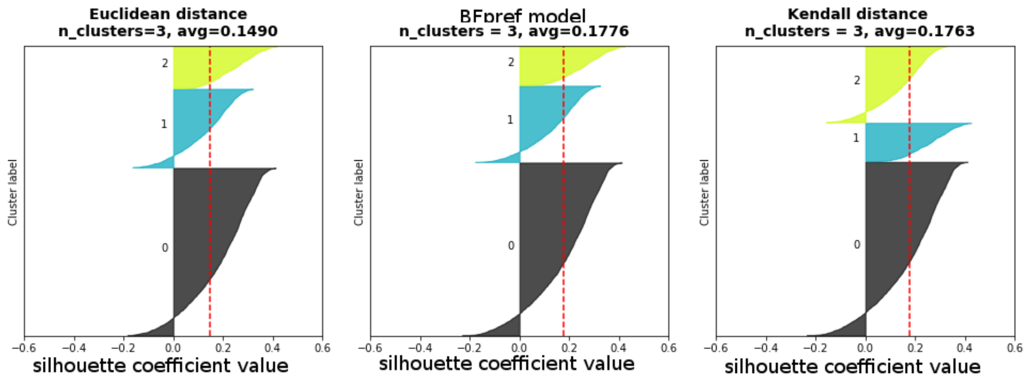


Figure 7.5 – Silhouette plot of different metrics on SUSHI dataset

- For least preferred two alternatives (2 alternatives at the end of the σ_r, i.e. ranking no. 9 and 10), we give score 1.
- For alternatives sorted at the positions 7 and 8, we give score 2.
- With the similar rule, for most preferred two sushis (ranking no. 1 and 2), we give score 5.

For example with: $a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6 \succ a_7 \succ a_8 \succ a_9 \succ a_{10}$
the scores are: $a_1 : 5, a_2 : 5, a_3 : 4, a_4 : 4, a_5 : 3, a_6 : 3, a_7 : 2, a_8 : 2, a_9 : 1, a_{10} : 1$

We compared our model with an average-based-euclidean metric. Still, ARI and silhouette scores are applied as evaluation criteria.

Confronting a case of two preference sources: ranking σ_r and score σ_s, the mean rank of alternative a_i is calculated:

$$\bar{r}(a_i) = \frac{1}{2}(r(\sigma_r, a_i) + r(\sigma_s, a_i)) \tag{7.23}$$

Thus, agent u's average preference order is represented by:

$$\bar{O}_u := \left[\bar{r}(a_1), \quad \bar{r}(a_2), \quad \ldots, \quad \bar{r}(a_{|A|})\right] \tag{7.24}$$

Therefore, the example above has vector:

$$\begin{bmatrix}1, & 1.5, & 3, & 3.5, & 5, & 5.5, & 7, & 7.5, & 9, & 9.5\end{bmatrix}$$

For Kendall distance, we calculate the distance matrix from rankings and scores, then take the average value as the combined distance. As indifference relation exists in σ_s, we apply Fagin distance for σ_s. Given ranking preferences σ_{r1}, σ_{r2} and score preferences σ_{s1}, σ_{s2}, denoting respectively the preference from agents agt_1 and agt_2, the average distance is thus given by:

$$\bar{d}_{Fagin}(agt_1, agt_2) = \frac{1}{2}(d_{Fagin}(\sigma_{r1}, \sigma_{r2}) + d_{Fagin}(\sigma_{s1}, \sigma_{s2})) \tag{7.25}$$

We compared the model based on Euclidean distance equation (7.24), Kendall distance (7.25) and BFpref model given in equation (7.8).

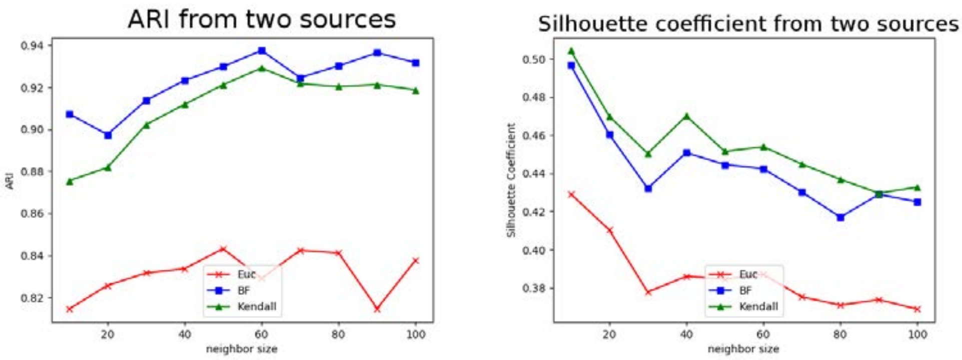


Figure 7.6 – ARI and silhouette coefficient on conflicting preferences, switch = 1

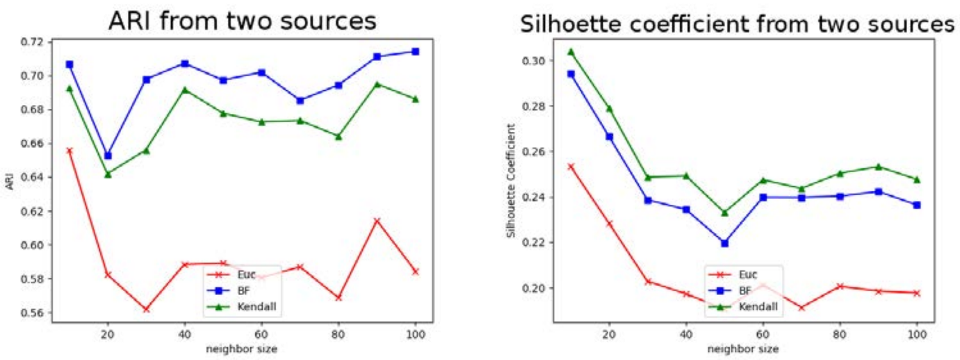


Figure 7.7 – ARI and silhouette coefficient on conflicting preferences, switch = 2

The results illustrated by Figures 7.6,7.7 and 7.8 show the advantage of BFpref model over Euclidean distance and Kendall (Fagin) distance when dealing with two

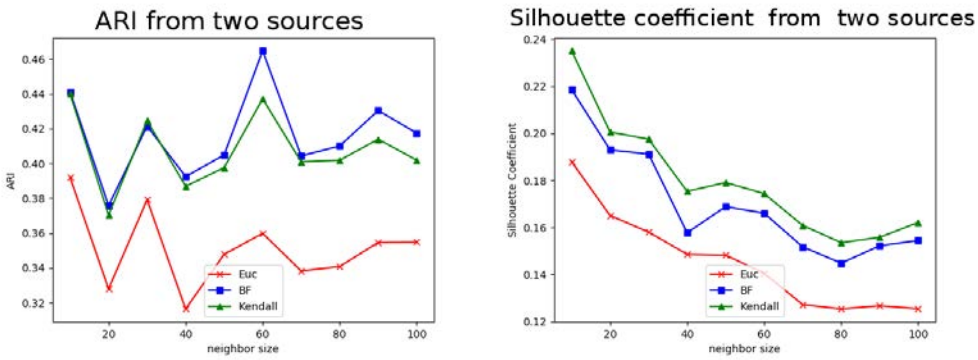


Figure 7.8 – ARI and silhouette coefficient on conflicting preferences, switch = 3

sources. Comparing Figures 7.6, 7.7, and 7.8 from conflicting sources with Figures 7.2, 7.3, and 7.4, we observe that both averaged Euclidean distance and Kendall distance are deteriorated more than BFpref model. The results prove the advantage of BFpref model on preferences under uncertainty. This advantage comes from the fact that in BFpref model, conflicts are partly interpreted as ignorance and have less impact in dissimilarity measuring. However, this compromise also causes a loss in criterion of silhouette coefficient.

7.5 Experimentation results for k determination in EkN-Nclus

In this section, the experiment as well as illustrative examples for k determination in EK-NNclus method is given. We firstly studied the correlation between ARI and silhouette coefficient, and then applied our strategy on toy datasets. The synthetic data are generated by Gaussian distributions. For the sake of better visualization, the synthetic data are always generated in a 2 dimensional space.

7.5.1 Correlation between ARI and silhouette coefficient

We generate synthetic datasets for this experiment. The procedure is as follows:

1. Given a set of standard deviation (noted *std*) and the number of clusters denoted by n_{clus}, we generate a set of datasets $\mathcal{S}_{data} = \{X_1, X_2, \ldots, X_D\}$ with ground truth. Datasets with 8 clusters and with std = 0.5, 1.0, 2, 2.5 are illustrated in Figure 7.10.

2. On one dataset $X_d \in \mathcal{S}_{data}$, given a set of parameter values $\mathcal{K} = \{k_1, k_2, \ldots, k_{|K|}\}$, calculate ARI and silhouette coefficient of each $k \in \mathcal{K}$. A set of ARIs and silhouette coefficients are obtained corresponding to different k, respectively denoted as $\mathcal{S}_{ARI}$ and $\mathcal{S}_{sil}$. The Pearson correlation coefficient $\rho(\mathcal{S}_{ARI}, \mathcal{S}_{sil})$ is calculated for dataset X_d, denoted by ρ_d.

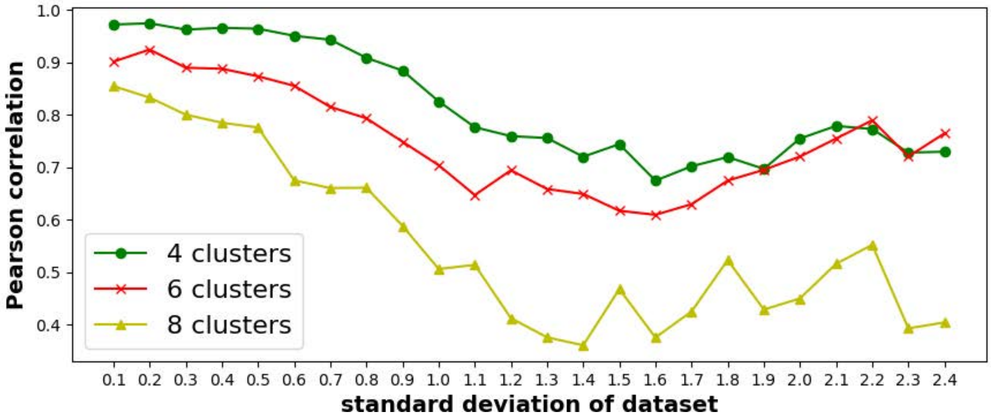


Figure 7.9 – Pearson correlation coefficient between ARI and silhouette ***vs.*** data sets with different *std*.

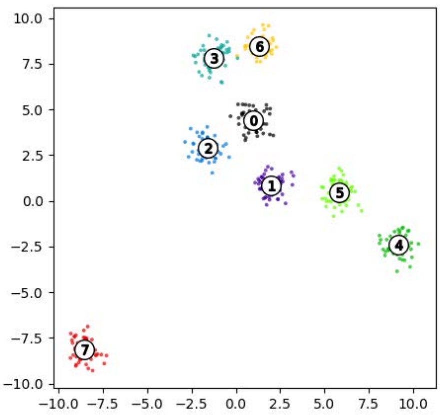
Figure 7.9 illustrates the variation of the correlation between ARI and silhouette coefficient via different standard deviations. We observe that the correlation declines while data are distributed more sparsely. From a certain standard deviation, the correlation has a tendency to increase. These are datasets used in the experiment of Figure 7.9. While *std* is small, data are obviously clustered. Thus a clustering result regrouping objects nearby is consistent with the knowledge of the ground truth, which returns a high correlation. With *std* increasing, different clusters overlap and the correlation decreases. When *std* is high enough that data distribution converges to random, the clustering returns lows values on both ARI and silhouette coefficient, making them "correlated" again. Of course, this is not a real correlation. Both ARI and silhouette are measured on random generated data, the results are always bad, making the correlation index high.

However, the strong correlation cannot guarantee that silhouette coefficient is enough for k determination. The ARI and silhouette coefficient obtained from different k on data in Figure 7.10 are respectively plotted in Figure 7.11. We observe that a high silhouette coefficient does not always correspond to a high ARI when value of k is large, even if objects in different clusters are naturally well separated (***e.g.*** dataset with *std* = 0.5). This has been explained in Section 7.3.1 that a high value on k may cause underestimation of the number of clusters c, which may result in a satisfying silhouette coefficient. Elbow method determining the c helps to provide a constraint condition.

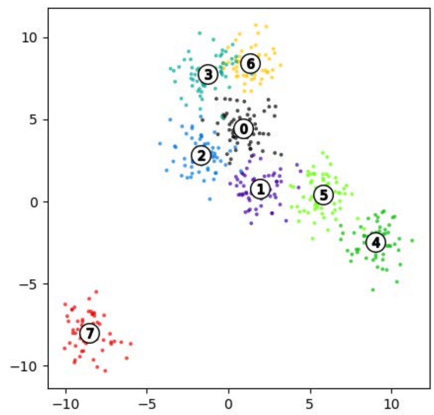
7.5.2 Optimal k determination strategy on real toy datasets

We applied the strategy in Section 7.3.3 on real toy datasets: Iris and Wine datasets from UCI[4] to help to refine the interval of k.

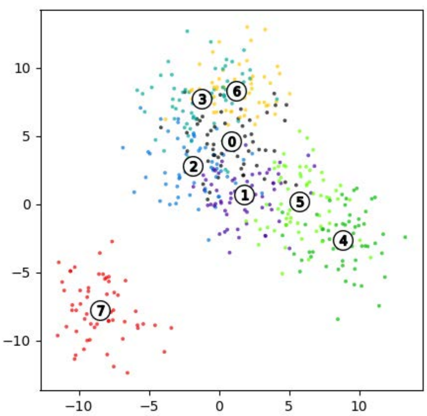
[4]Iris: https://archive.ics.uci.edu/ml/datasets/Iris
Wine: https://archive.ics.uci.edu/ml/datasets/wine



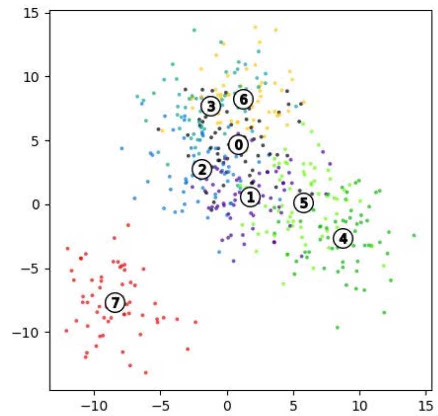
(a) std=0.5



(b) std=1.0



(c) std=2.0

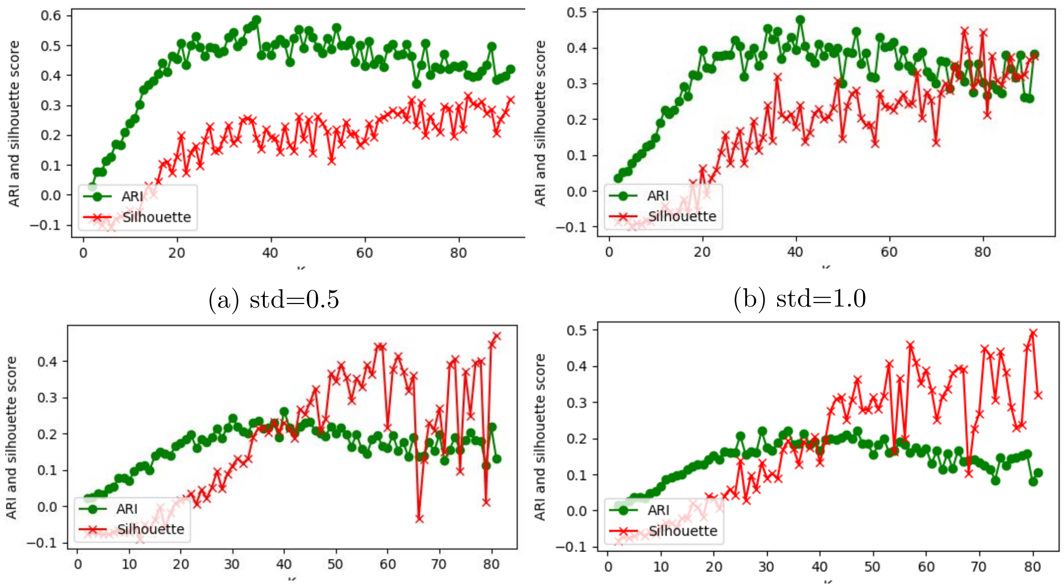


(d) std=2.4

Figure 7.10 – Data distributions with different values of standard deviation.

Toy dataset *Iris*: Figure 7.12 illustrates the plot supporting k determination strategy for Iris toy data. Results are obtained with a cross validation of 10 experiments. We still observe that the values of ARI, silhouette coefficient and number of clusters have large fluctuation, which proves that the determination of k is risky.

Without knowledge of c, from the silhouette coefficient plot in Figure 7.12b, one may conclude that $k \in [30, 50]$ is the best value. With elbow method, we can figure that $c = 2$ or 3 is a reasonable value, so $k \in [15, 40]$ is more reasonable. Taking the intersection of both intervals, we focus on a refined interval $k \in [30, 40]$. In this interval, $k = 35$ returns the highest silhouette coefficient (given by the abscissa of Figure 7.12b). Thus, finally we determine $k = 32$ by equation (7.22). With the ARI plot (given by the ordinate of Figure 7.12b), we can verify that $k \approx 35$ is the proper value, so the



(a) std=0.5 (b) std=1.0

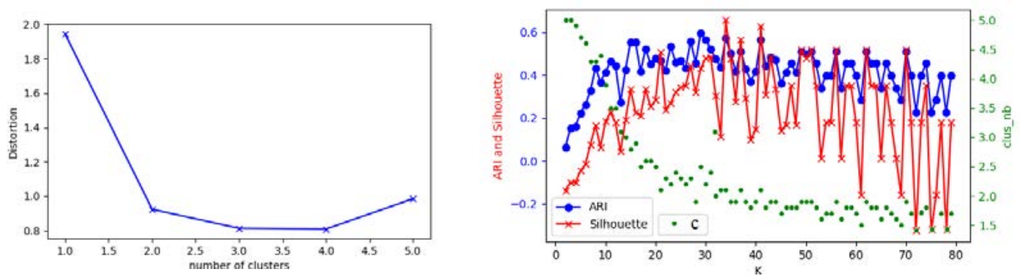
(c) std=2.0 (d) std=2.4

Figure 7.11 – ARI and Silhouette coefficient *via* k on different datasets.

proposed strategy is adapted.

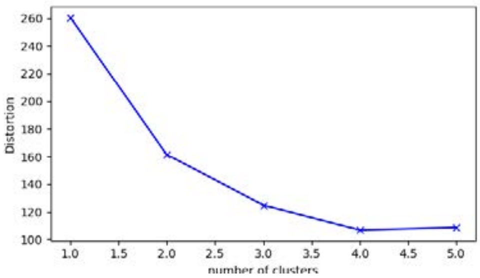
Toy dataset *Wine*: The elbow method and clustering criteria plot are illustrated in Figure 7.13. It is tricky to determine the number c of clusters by Elbow method for this dataset. Different observers may give different decisions on the best number of clusters. Therefore, 3 or 4 can both be concluded as c. According to Figure 7.13b, $c \in \{3, 4\}$ corresponds approximately to $k \in [20, 50]$. A high silhouette coefficient value corresponds to the interval $k \in [40, 70]$. By taking the intersection of both intervals, we conclude that a proper k should be in the interval $[40, 50]$ and we obtain $k = 49$ such as the optimal value by equation (7.22).

According to Figure 7.13b, with only silhouette coefficient, we may arbitrarily

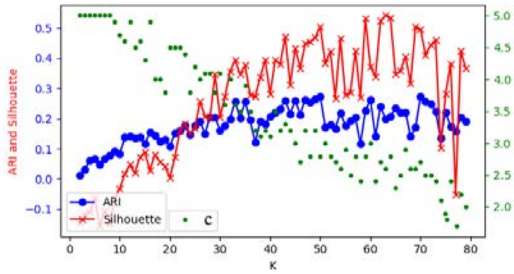


(a) Elbow method. (b) ARI, silhouette coefficient and number of clusters *vs.* k.

Figure 7.12 – Results on Iris dataset.



(a) Elbow method.



(b) ARI, silhouette coefficient and number of clusters *vs.* k.

Figure 7.13 – Results on Wine dataset.

choose a high value $k \in [60, 70]$. However, this value gives an underestimation of the c value. The elbow method fixing a proper number of clusters helps to determine a k that returns the highest ARI.

7.6 Conclusion

In this chapter, we investigate the problem of clustering agents according to their preferences, when dealing with multiple and conflicting sources (two in our case study). To cope with this issue, we apply the BFpref model to express and interpret the contradictions and conflicts from different sources as uncertainty and ignorance. We introduce a new approach that captures the preference data structure and deal with uncertain information.

To highlight the relevance of the proposed solution, we perform experiments on synthetic and real data to compare our method with other preference models, and found the advantage in the expressiveness of the uncertainty and the incomparability of the preference orders. Indeed, we compare BFpref model on synthetic data between Euclidean distance and Kendall distance both in certain and uncertain cases, using EK-NNclus algorithm. In certain cases, BFpref model has equivalent clustering-quality with Kendall distance and outperforms the Euclidean distance. In uncertain cases, BFpref model has better clustering-quality over the other distances. BFpref model is also applied on SUSHI preference data set and observed that BFpref model has one of the most satisfying clustering-quality.

BFpref model is only applied on complete preference orders (***i.e.*** weak orders) from only two sources. In the future, we will work on an ameliorated BFpref model dealing with several conflicting preferences sources. A more general dissimilarity measure method for incomplete orders (***i.e.*** quasi-orders) is also in the scope of our future work.

Besides, a practical problem encountered in the application of EK-NNclus algorithm is also discussed: the determination of the optimal number of nearest neighbors k. Based on some methods borrowed from determination of the number c of clusters in c-means, we proposed a combined strategy. In this strategy, silhouette coefficient is applied to evaluate the clustering quality and elbow method is used as an extensive procedure for over-fitting. Comparing with an empirical suggestive interval for k determination given by [DKS15b], the proposed strategy gives a more refined selection of k and guarantees a relative high quality of clustering.

The strategy has some short-comings conducted by elbow method. Firstly, the determination of c by elbow method is subjective and can be sometimes ambiguous. Besides, the distortion requires the calculation of centroïds of clusters, which neutralizes an advantage of EK-NNclus: EK-NNclus is centroïd independent. This issue leaves a pity that the k determination method is not applicable in clustering on evidential preferences. In the future, we plan to replace elbow method by centroïd-independent c determination method, making the strategy more adaptable.

Conclusion and perspectives

Conclusion

Preference has been a popular topic in scientific research for long time ranges from sociology, economy to computer science and engineering. In this thesis, we focus on the modeling and management on imperfect preference information, applied in decision making and machine learning.
Three aspects of imperfection on preference information are considered: uncertainty, imprecision and incompleteness. In preference information, the uncertainty is a situation of limited knowledge where it is impossible to exactly describe the existing preference relation. This is usually an epistemic situation. The imprecision refers to situations where multiple preference relations are possible, so that it is impossible to decide on a specific relation. Preferences with uncertainty and imprecision is also called “evidential preferences”. The incompleteness in preferences considered in the thesis refers to cases where preference relations are unknown between some alternatives, making it impossible to create complete preference relations on all alternatives. The imperfections are usually caused by the flaws in the sources of preference data, such as uncertain opinion of agents, conflicts in multiple sources and implicit information. The research topic of this thesis is a combined domain of preference, machine learning imperfect data modeling.

In the state-of-the-art part, we firstly reviewed different modeling methods on both crisp and uncertain preference as well as the theory of belief functions. Total orders, weak orders, partial orders and quasi orders are compared in the definition level. Among many uncertain preference models, we highlighted fuzzy preference, which is the most popular one for uncertain preference modeling. However, fuzzy preference is capable to express the uncertainty but not imprecision problems. In the part of the theory of belief functions, besides the basic concepts, some commonly used combination rules are also introduced with adaptable application circumstances.

After this, we introduce on similarity measuring methods on preference relations and structures as well as similarity measure in the theory of belief functions. Between two preference relations, in apart to different norms of distances, the axiomatic distances are specific on preferences. Different axiomatic preferences are compared from the level of axioms accepted to values adopted. As for similarity between preference structures, different correlation based distances are introduced, such as Kendall distance, Spearman footrule distance and Peason correlation.

Afterwards, different preference management problems and solutions, more specifically, aggregation and preference clustering methods, are introduced as well as the role that similarity plays in these processes. In the preference aggregation part, typical voting rules in election systems as well as different aggregation strategy for fuzzy preference are introduced, followed by traditional problems such as Condorcet paradox and Arrow's impossibility theorem. In preference clustering part, we clarify the position of preference clustering in the research topic of Artificial Intelligence. Confronting the incompleteness problem in machine learning, we categorize the mainstream methods into three groups: data discarding strategy, data imputation strategy and soft method strategy.

The contributions are also on the modeling and management on imperfect preference data. Firstly, we pointed out an ambiguity in the "incompleteness" caused by the definition of preference relation "incomparability". Most work interpret "incomparability" as a missing information case, or "not decided" case. Some other works interpret "incomparability" as a specific binary relation that is different from "strict preference" and "indifference", or "non-decisive" case, which also respect the original definition of "incomparability",. In this work, we consider the incompleteness in the preferences is caused by the missing information and clarified this ambiguity in an novel evidential preference model based on the theory of belief functions, namely BFpref model. BFpref model is a pairwise preference model with the frame of discernment made up by four singleton, respectively representing "strict preference", "inverse strict preference", "indifference", and "incomparability". BFpref model is capable to express all the three aspects of imperfection in preferences, and it distinguishes the ambiguity in the definition of "incomparability". The "non-decisive" case is directly represented by a singleton while the missing information (not-decided case) is represented by an extreme imprecision case–total ignorance, which is the union off all possible singleton in the frame of discernment.

With the help of BFpref model, we proposed an aggregation strategy on evidential preferences based on Dempster's combination rule and minimum distance decision strategy. We also proposed a strategy to avoid Condorcet paradox as well as an efficient DFS method for Directed Acyclic Graph building.

During the decision step of preference aggregation on BFpref model, we pointed out a flaw in state-of-the-art distances for evidential objects. Some distances in TBF are structural, ***i.e.*** they account for the interaction between the focal elements compared in two BBAs. However, none of them is capable to distinguish the measure on BBAs of singletons with different weights.
To solve this problem, we analyzed assumptions accepted by Jousselme distance and dropped an unwanted one, which consider the different singletons are equally different.

By extending Jousselme distance and proposed WSD distance. We also applied WSD in decision making on SUSHI data set. The comparison results demonstrated that decisions based on WSD is more reasonable than on Jousselme distances.

Learning on evidential preference, more specifically, clustering on preferences in BFpref model, is also in the scope of this thesis. We tended to represent agents' portraits by evidential preference by BFpref when multiple sources of preference with conflicts are considered. The representation works are mostly in estimation step. The similarity between agents are based on the sum of Jousselme distance and degrade to Kendall distance in case of crisp preference. With comparison with a mean value strategy, it is illustrated that BFpref returns better clustering result in terms of silhouette score confronting multiple preference sources with conflicts for identical agents. However, the clustering work is solely done on complete preferences. For the incomplete evidential preferences, we proved an impossibility theorem that there is no pertinent combination rule for centroid calculating among evidential objects, which respects the properties of "metric consistency", "uniqueness of combination" and "ignorance neutrality". The research on evidential preferences learning can be generalized into learning on evidential objects, which is in the perspective of this thesis. Several new scientific issues are proposed in the work of this thesis. Some of them are our future work introduced as perspectives.

Perspectives

In this section, we discuss on some scientific challenges related to the topic of this thesis, from aspects of short and long terms perspectives. The introduction of perspectives starts with a discussion on similarity measure between evidential objects.

More discussion on similarity measure between evidential objects

In Section 7.1, we proposed an impossibility theorem on distance between evidential object in clustering. In this section, we give more discussion on this topic.[5]

For learning on objects expressed on BBAs, we believe that a conflict based method consistent with Dempater's conjunctive combination rule $\textcircled{\cap}$ may be more pertinent. One possible solution is given as follows:
Between two BBAs m_1 and m_2, instead of a distance function $d : \mathcal{E}(\Omega) \times \mathcal{E}(\Omega) \to \mathbb{R}_{\geq 0}$, a measure function for evidential objects d_{TBF} should return an evidential value, ***i.e.*** The similarity measure is essentially a projection function in from the original discernment to a new one. Formally, for a set of evidential objects BBAs

$$d_{TBF} : \mathcal{E}(\Omega) \times \mathcal{E}(\Omega) \to \mathcal{E}(\Omega') \tag{7.26}$$

[5] Acknowledgment:
The main idea of this discussion is inspired by a discussion with Pr. Sébastien Destercke from UTC.

where Ω' is another discernment frame than original discernment Ω for similarity measure.
For m_1, m_2, the simplest discernment frame for clustering Ω' can be defined as:

$$\Omega' = \{\omega'_1, \omega'_2\} \tag{7.27}$$

where ω'_1 indicates that m_1 and m_2 are in the same cluster while ω'_2 indicates that the two BBAs are in different clusters. With such definitions, the clustering procedures with multiple feature dimensions $\mathcal{D} = \{m_{(1)}, m_{(2)}, \dots m_{(D)}\}$, a method with following steps may be feasible:

1. Define a measure function $d_{TBF}^{dim}(\cdot)$ for each dimension $dim \in \mathcal{D}$. For instances i and j, get the estimated partition BBA $m_{i,j,(dim)}$ (Estimation step)

2. For instances i, j, combine on all feature dimensions with conjunctive rule (Combination step):

$$m_{ij} = \bigcap_{dim \in \mathcal{D}} m_{i,j,(dim)} \tag{7.28}$$

   The combination step may be discounting coefficient defined in Equation (2.26), repeated here:

$$m_j^{\alpha}(X) = \alpha_j m_j(X), \forall X \subsetneq \Omega$$
$$m_j^{\alpha}(\Omega) = 1 - \alpha_j(1 - m_j(\Omega))$$

   with α_j estimated by learning or other methods.

3. Partition evidential objects based on $m_{i,j}$

In short term, we will continue the research work on preference learning on evidential preferences with incompleteness. Confronting this challenge, several issues need to be resolved.

Short term: preference learning with BFpref model

In this thesis, we have studied the clustering problem on complete evidential preference with BFpref model. The objective is to develop a method for predicting the unobserved preference relations on incomplete evidential preference. However, we have proven some difficulties in measuring the distance between BBAs especially when vacuous BBAs exist, representing unobserved preferences. To address this issue, following problems should be firstly responded:

What does similarity really measure between evidential objects?

When measuring the similarity between evidential preferences, we actually want to measure the similarity between the objects represented by BBAs rather than BBA vectors. Thus, the similarity between BBAs does not always correspond to the similarity between evidential objects. An extreme example has been discussed that two identical vacuous BBAs do not indicate that the evidential pieces of knowledge are identical.

Indeed, a proper result should be that the similarity between two vacuous BBAs is also total ignorance, i.e. it's impossible to measure such similarity.
All the state of the art distances for BBAs measure either the vector representation or the divergence between uncertainty. The first one measures the similarity from a geometry view and the latter from a statistic view. However, none of them measures the objects themselves. We believe that the reason of this issue is that the one single BBA may express multiple kinds of information. Here is an example, a categorical BBA describes purely the state of the an object, or "content" (the state is a qualitative value) While a vacuous BBA describes purely the evidence degree (uncertainty and imprecision) which is total ignorance. Other BBAs may express these two kind of information at the same time.
To clarify this issue, we reckon that the concept of information quantity of a BBA may be helpful, which is usually measured by entropy. In [JS18], they authors made a survey on properties in different definitions of entropy in the theory of belief functions. A well defined similarity measure for BBAs should take account of both "content" and "evidence degree". As discussed in Section 7.1.4, it is more pertinent to measure the similarity between BBAs by a BBA in another discernment framework (We call it ***evidential similarity***).

Thus, the measure of similarity on objects represented by BBAs is a projection process between two different frameworks of discernment. To realize such projection, proper properties and conditions should be firstly defined. This is the nearest work in the future.

How to apply evidential similarity in learning algorithms?

This question is one step further than the previous one. All the distance based machine learning algorithms apply "traditional distance functions" in non-negative real numbers. Assume that a proper evidential similarity is defined, corresponding machine learning algorithms need to be designed. Thus, to develop a such method is also another important perspective. As similarities are no longer crisp, soft methods in machine learning may be helpful.

Long term: more applications with BFpref model

The long-term perspectives concerns more applications with BFpref model. We propose several potential applications and issues to address.

How to manage imperfect preferences in a multi-criteria context?

In the problems adressed in this thesis, alternatives are represented by a single criterion, thus these methods are also named as mono-criterion decision making methods. In some circumstancs, alternatives are represented by multiple criteria, proposing multi-criteria decision making problems (MCDM). Indeed, mono-criterion decision making is the base of multi-criteria decision.

There are already various of group decision making methods developed on MDCM problems, such as analytic hierarchy process (AHP) developed by Thomas L. Saaty in 1970s [Saa02], ELECTRE ("ELimination Et Choix Traduisant la REalité" or "ELimination and Choice Expressing REality") [Roy68, FMR05], PROMETHEE [BV85, BM05]. In multi-criteria context, preferences may also be imperfect. For uncertain cases, fuzzy set approaches: [KF87, YDP99], rough set approaches [GMS99a, GMS99b], optimization approaches [LS02, CK91] and probabilistic models [Ste05] are also important works.

In the future, we well study more problems of imperfectness in multi-criteria context, such as imprecision and incomplete problems. We believe that BFpref model can be useful in multi-criteria decision making problems as well as multi-criteria preference learning problems confronting preferences with uncertainty, imprecision as well incompleteness.

How to apply BFpref on data of large scale?

In the theory of belief functions, for a value of $|\Omega|$ states, the discernment frame is defined on a space with dimension of 2^{Ω}. In BFpref, for one pair of alternatives, 16 values is registered. Moreover, as BFpref model is based on pairwise relation, on N alternatives, its space complexity is $\mathcal{O}(N^2)$, while list-wise order model has complexity of $\mathcal{O}(N)$. When the alternative space scales up, the require of memory space increase rapidly, making the calculation costly. In order to apply BFpref in industry circumstances, it is important to ameliorate the corresponding process.

This work can be two folded. First idea is in the expression of BBAs and second in the calculation process.
Although each BBA is expressed in space of 2^{Ω}, most of elements are valued as 0. Thus, it is sufficient to assign only non-zero elements and ignore the elements with 0. For example, a simple BBA only needs to express one non-zero element.
The calculation of distances between preferences on different alternative pairs are independent. This makes it possible to propose algorithms for parallel or distributed computing, which is practical in dealing with big scale of data.

How to apply BFpref on real world data

In the world of internet, there are usually different websites or services who work on similar business (***i.e.*** they are competitors). ***eg.*** Yelp and TripAdviser provide restaurant recommendation service, Amazon and eBay provide retail service, Booking and PriceLine provide room and vehicle reservation services, Rotten Tomatoes and IMDb provide film rating and reviewing services. Therefore, for consumers, it is common to have accounts on different sites in the same field. This make it possible that one individual's preference on identical alternatives may have different sources. (For example, one may rate Avengers III higher than Captain American III on IMDb while equally on Rotten Tomatoes). This is similar to the case we encountered in SUSHI data. We believe that it is valuable to build a reliable and robust recommendation system based on preferences from multiple sources. Furthermore, we would also like to apply BFpref model in social network analysis, especially in community detection. In this perspective, an important challenge exist in the estimation of BBAs from different

Appendices

A Proof of proposition 1

Proposition 1:

In the frame of discernment $\Omega = \{\omega_1, \ldots, \omega_k\}$. If for a BBA m, exist ω_d such that

$$\forall \omega_i \in \Omega, \omega_i \neq \omega_d,$$

$$betP_m(\omega_d) \geqslant betP_m(\omega_i)$$

Proof. We rewrite the Jousselme distance by its square form for simplification. Thus,

$$\begin{aligned}
& d_J^2(\boldsymbol{m}, \boldsymbol{m}_{cat}^{\omega_d}) \\
=& (\boldsymbol{m} - \boldsymbol{m}_{cat}^{\omega_d})^T \boldsymbol{Jacc} (\boldsymbol{m} - \boldsymbol{m}_{cat}^{\omega_d}) \\
=& \sum_{\substack{X \neq \omega_d \\ Y \neq \omega_d}} m(X)m(Y)Jacc_{X,Y} \\
& + \sum_{\omega_d \in Y} (m(\omega_d) - 1)m(Y)\frac{1}{|Y|} + \sum_{\omega_d \in X} (m(\omega_d) - 1)m(X)\frac{1}{|X|} \\
=& \sum_{\substack{X \neq \omega_d \\ Y \neq \omega_d}} m(X)m(Y)Jacc_{X,Y} \\
& + \sum_{\omega_d \in Y} m(\omega_d)m(Y)\frac{1}{|Y|} + \sum_{\omega_d \in X} m(\omega_d)m(X)\frac{1}{|X|} \\
& - \sum_{\omega_d \in Y} m(Y)\frac{1}{|Y|} - \sum_{\omega_d \in X} m(X)\frac{1}{|X|} \\
=& \sum_{\substack{X \neq \omega_d \\ Y \neq \omega_d}} m(X)m(Y)Jacc_{X,Y} \\
& + \sum_{\substack{X = \omega_d \\ \omega_d \in Y}} m(X)m(Y)Jacc_{X,Y} + \sum_{\substack{\omega_d \in X \\ Y = \omega_d}} m(Y)m(X)Jacc_{Y,X} \\
& + \sum_{\substack{X = \omega_d \\ \omega_d \notin Y}} m(X)m(Y)Jacc_{X,Y} + \sum_{\substack{\omega_d \notin X \\ Y = \omega_d}} m(Y)m(X)Jacc_{Y,X} \\
& - \sum_{\omega_d \in Y} m(Y)\frac{1}{|Y|} - \sum_{\omega_d \in X} m(X)\frac{1}{|X|} \\
=& \sum m(X)m(Y)Jacc_{X,Y} - 2\sum_{\omega_d \in Y} m(Y)\frac{1}{|Y|}
\end{aligned} \tag{29}$$

Similarly, we have

$$d_J^2(m, m_{cat}^{\omega_i}) = \sum m(X)m(Y)Jacc_{X,Y} - 2\sum_{\omega_i \in Y} m(Y)\frac{1}{|Y|} \tag{30}$$

Given that

$$betP_m(\omega_d) \geqslant betP_m(\omega_i)$$

we have

$$\frac{1}{1-m(\emptyset)}\sum_{\omega_d\in Y} m(Y)\frac{1}{|Y|} \geqslant \frac{1}{1-m(\emptyset)}\sum_{\omega_i\in Y} m(Y)\frac{1}{|Y|}$$

$$\Rightarrow \sum_{\omega_i\in Y} m(Y)\frac{1}{|Y|} - \sum_{\omega_d\in Y} m(Y)\frac{1}{|Y|} \leqslant 0$$

Therefore

$$\begin{aligned} & d_J^2(m, m_{cat}^{\omega_d}) - d_J^2(m, m_{cat}^{\omega_i}) \\ = & -2\sum_{\omega_d\in Y} m(Y)\frac{1}{|Y|} + 2\sum_{\omega_i\in Y} m(Y)\frac{1}{|Y|} \leqslant 0 \\ & \Rightarrow d_J(m, m_{cat}^{\omega_d}) \leqslant d_J(m, m_{cat}^{\omega_i}) \end{aligned}$$

Inequation (6.5) is true.

□

B Proof of proposition 2

Proposition 2: Let a combination rule $\odot$ with the property of consistent metric for a distance d, then the combination rule $\odot$ is idempotent.

Proof. Let a combination rule $\odot$ with the property of consistent metric for a distance d and let m a BBA and $\forall$ BBA $m' \neq m$, according to the property of identity of indiscernible and the non-negativity of the distance d, we have:

$$d(m, m') > 0.$$

with $d(m, m) = 0$,

$$d(m, m) + d(m, m) < d(m, m') + d(m, m')$$

Thus,

$$m \odot m = m$$

□

Proposition 2: There is no combination rule owning properties of "metric consistence", "uniqueness of combination result" and "Ignorance neutrality".

Proof. Given a BBA $m \neq m_\Omega$, according to the property of idempotent, we have

$$\forall m' \neq m$$

$$d(m, m) + d(m, m) \leqslant d(m, m') + d(m, m')$$

with $d(m, m) = 0$ and $m \neq m'$ (identity of indiscernible), we have

$$d(m, m') > 0. \tag{31}$$

(This proves that the idempotent is an intrinsic property of metric consistence)

With the uniqueness of combination result, in relation (7.5), the equality becomes impossible. Thus, Inequality (7.5) becomes

$$\sum_{i=1}^{k} d(m_k, m') > \sum_{i=1}^{k} d(m_k, m_{comb}) \tag{32}$$

With the property of ignorance neutrality, from the relation inequality (32), we have $\forall m' \neq m$,

$$d(m, m) + d(m, m_\Omega) < d(m, m') + d(m', m_\Omega) \tag{33}$$

Take $m' = m_\Omega$, Inequations (31) and (33) become

$$d(m, m_\Omega) > 0 \tag{34}$$

and

$$d(m, m_\Omega) < d(m, m_\Omega) + d(m_\Omega, m_\Omega)$$

written as:

$$d(m_\Omega, m_\Omega) > 0$$

This is contradictory to the property of identity discernible. Thus the assumption is false. □

C Proof of proposition 3

Proposition 3: Based on Jousselme distance, we define a normalized distance for evidential preference orders σ_1 and σ_2:

$$d_{Normalize}(\sigma_1, \sigma_2) = \frac{1}{Nb_{total}} d(\sigma_1, \sigma_2)$$

$d_{Normalize}(\sigma_1, \sigma_2)$ is equivalent to Kendall distance when the preference orders are total and crisp.

Proof. Kendall distance: Given two complete ranking orders σ_1 and σ_2, the Kendall τ distance d_τ is defined as:

$$d_{Kendall}(\sigma_1, \sigma_2) = \sum_{rank_{\sigma_1}(a_i) < rank_{\sigma_1}(a_j)} (rank_{\sigma_2}(a_i) > rank_{\sigma_2}(a_j)), \tag{35}$$

This is equivalent to the following equation:

$$d_{Kendall}(\sigma_1, \sigma_2) = \sum_{i<j} \bar{K}_{i,j}(\sigma_1, \sigma_2) \tag{36}$$

where

- $\bar{K}_{i,j}(\sigma_1, \sigma_2) = 0$ if i and j are in the same relative ranking in σ_1 and σ_2
- $\bar{K}_{i,j}(\sigma_1, \sigma_2) = 1$ if i and j are in the opposite order in σ_1 and σ_2.

In BFpref model, denote the preference a_i and a_j in σ_1 and σ_2 as $m_{ij,1}$ and $m_{ij,2}$. If the preference is crisp, the two BBAs are categorical. If a_i and a_j are in same relative rank in both σ_1 and σ_2, we have

$$d_J(m_{ij,\sigma_1}, m_{ij,\sigma_2}) = 0$$

Otherwise, a_i and a_j are in different relative rank in σ_1 and σ_2 (let's say $a_i \succ a_j$ in σ_1 and $a_i \prec aa_j$ in σ_2), we have

$$m_{ij,\sigma_1}(\omega_{\succ}) = 1$$

$$m_{ij,\sigma_2}(\omega_{\prec}) = 1$$

the Jousselme distance between m_{ij,O_1} and m_{ij,O_2}

$$d_J(m_{ij,\sigma_1}, m_{ij,\sigma_2}) = 1$$

For a preference structure, Jousselme distance based similarity is defined as:

$$d_{J,structure}(\sigma_1, \sigma_2) = \sum_{i<j} d_J(m_{ij,\sigma_1}, m_{ij,\sigma_2}) \tag{37}$$

This is equivalent to Equation (36). Similarly, the proposition is also true if Jousselme distance is replaced by WSD. □

D Aggregation result of sushi preference in east and west Japan

The ranking results of east Japan and west Japan under two different distance for decision making are illustrated below, with numbers 0 to 99 indicating the types of sushi.

- Aggregation of East Japan by decision rule of Jousselme distance:
  19, 8, 37, 9, 2, 15, 61, 6, 1, 41, 47, 11, 79, 53, 0, 10, 22, 20, 13, 26, 27, 4, 3, 21, 31, 36, 76, 25, 5, 57, 34, 14, 7, 73, 46, 54, 32, 45, 43, 38, 65, 50, 88, 62, 18, 44, 58, 82, 12, 66, 72, 63, 71, 39, 70, 23, 91, 16, 48, 51, 35, 33, 67, 29, 60, 96, 24, 74, 28, 75, 68, 30, 77, 83, 42, 95, 55, 80, 52, 64, 92, 56, 85, 49, 69, 78, 87, 59, 99, 84, 40, 90, 81, 86, 17, 98, 93, 89, 94, 97

- Aggregation of West Japan by decision rule of Jousselme distance:
  8, 19, 15, 9, 11, 2, 0, 13, 53, 20, 47, 61, 1, 22, 6, 3, 37, 26, 4, 41, 27, 10, 79, 43, 14, 5, 21, 36, 34, 25, 45, 65, 88, 31, 73, 57, 30, 7, 76, 46, 39, 44, 12, 68, 48, 71, 62, 54, 67, 50, 70, 72, 55, 32, 74, 18, 24, 95, 66, 82, 96, 52, 33, 63, 23, 51, 83, 29, 35, 87, 58, 60, 49, 28, 75, 64, 85, 78, 16, 84, 69, 40, 81, 38, 56, 80, 89, 77, 94, 17, 90, 99, 91, 42, 86, 59, 93, 98, 92, 97

- Aggregation of East Japan by decision rule of WSD distance
  8, 19, 2, 9, 37, 6, 15, 47, 61, 1, 11, 53, 0, 41, 22, 10, 79, 20, 13, 26, 4, 27, 3, 21, 31, 36, 25, 76, 5, 57, 34, 7, 14, 73, 54, 46, 43, 45, 32, 38, 65, 50, 88, 62, 18, 44, 72, 58, 66, 39, 71, 82, 23, 12, 63, 91, 70, 48, 51, 16, 35, 29, 60, 33, 75, 67, 74, 96, 24, 68, 95, 28, 30, 83, 80, 77, 42, 55, 64, 78, 52, 56, 85, 92, 99, 69, 49, 87, 84, 59, 40, 81, 90, 86, 17, 98, 93, 89, 94, 97
- Aggregation of West Japan by decision rule of WSD distance
  8, 19, 15, 9, 11, 2, 53, 47, 13, 20, 1, 0, 61, 3, 37, 6, 26, 22, 27, 4, 41, 10, 79, 43, 21, 14, 5, 36, 25, 34, 45, 65, 88, 31, 73, 76, 7, 46, 57, 48, 44, 71, 62, 68, 39, 30, 12, 50, 54, 67, 72, 70, 74, 18, 55, 32, 95, 66, 24, 82, 63, 96, 33, 52, 51, 83, 23, 29, 49, 60, 58, 87, 35, 75, 28, 85, 64, 78, 16, 84, 69, 81, 40, 56, 38, 80, 89, 17, 77, 94, 90, 99, 91, 42, 59, 86, 93, 98, 92, 97

E Sushi Information

Table 1 – Sushi information

ID	Sushi	original oiliness (0: most oily 4: least oily)	inverse oiliness (0: least oily 4: most oily)
0	ebi	2.73	1.27
1	anago	0.93	3.07
2	maguro	1.77	2.23
3	ika	2.69	1.31
4	uni	0.81	3.19
5	tako	3.09	0.91
6	ikura	1.26	2.74
7	tamago	2.37	1.63
8	toro	0.55	3.45
9	amaebi	1.91	2.09
10	hotategai	2.35	1.65
11	tai	2.99	1.01
12	akagai	2.52	1.48
13	hamachi	1.26	2.74
14	awabi	2.53	1.47
15	samon	1.27	2.73
16	kazunoko	2.68	1.32
17	shako	2.57	1.43
18	saba	1.92	2.08
19	chu toro	0.80	3.20
20	hirame	2.99	1.01
21	aji	2.32	1.68

22	kani	2.82	1.18
23	kohada	2.70	1.30
24	torigai	2.54	1.46
25	unagi	0.54	3.46
26	tekka maki	2.25	1.75
27	kanpachi	1.58	2.42
28	mirugai	2.39	1.61
29	kappa maki	3.73	0.27
30	geso	2.33	1.67
31	katsuo	1.91	2.09
32	iwashi	2.11	1.89
33	hokkigai	2.57	1.43
34	shimaaji	2.23	1.77
35	kanimiso	0.91	3.09
36	engawa	2.20	1.80
37	negi toro	1.26	2.74
38	nattou maki	1.65	2.35
39	sayori	2.80	1.20
40	takuwan maki	3.33	0.67
41	botanebi	1.99	2.01
42	tobiko	2.25	1.75
43	inari	1.58	2.42
44	mentaiko	1.81	2.19
45	sarada	2.48	1.52
46	suzuki	2.65	1.35
47	tarabagani	2.68	1.32
48	ume shiso maki	3.38	0.62
49	komochi konbu	2.24	1.76
50	tarako	1.88	2.12
51	sazae	2.41	1.59
52	aoyagi	2.50	1.50
53	toro samon	0.69	3.31
54	sanma	1.81	2.19
55	hamo	2.57	1.43
56	nasu	3.18	0.82
57	shirauo	3.08	0.92
58	nattou	1.53	2.47
59	ankimo	0.75	3.25
60	kanpyo maki	2.82	1.18
61	negi toro maki	1.26	2.74
62	gyusashi	1.15	2.85

63	hamaguri	2.29	1.71
64	basashi	1.48	2.52
65	fugu	3.13	0.87
66	tsubugai	2.40	1.60
67	ana kyu maki	2.22	1.78
68	hiragai	2.50	1.50
69	okura	2.37	1.63
70	ume maki	3.30	0.70
71	sarada maki	2.09	1.91
72	mentaiko maki	1.87	2.13
73	buri	0.96	3.04
74	shiso maki	3.50	0.50
75	ika nattou	1.58	2.42
76	zuke	1.49	2.51
77	himo	2.26	1.74
78	kaiware	3.41	0.59
79	kurumaebi	2.42	1.58
80	mekabu	3.05	0.95
81	kue	1.97	2.03
82	sawara	2.38	1.62
83	sasami	2.80	1.20
84	kujira	1.26	2.74
85	kamo	1.29	2.71
86	himo kyu maki	2.71	1.29
87	tobiuo	2.49	1.51
88	ishigakidai	2.66	1.34
89	mamakari	2.35	1.65
90	hoya	2.24	1.76
91	battera	1.68	2.33
92	kyabia	1.33	2.67
93	karasumi	1.71	2.29
94	uni kurage	1.44	2.56
95	karei	2.60	1.40
96	hiramasa	1.97	2.03
97	namako	1.94	2.06
98	shishamo	2.16	1.84
99	kaki	1.78	2.22

Remark: Original oiliness $oil_{original}$ is given from 0 to 4 indicating the most oily to the least oily. In our experiment, we use the data in ***inverse oiliness***, calculated by $4 - oil_{original}$ (in column ***inverse oiliness***) to correspond that a larger value indicate a more oily sushi.

Bibliography

[ACH+08] Sergio Alonso, Francisco Chiclana, Francisco Herrera, Enrique Herrera-Viedma, Jesús Alcalá-Fdez, and Carlos Porcel. A consistency-based procedure to estimate missing pairwise preference values. International Journal of Intelligent Systems, 23(2):155–175, 2008.

[AHVCH10] S. Alonso, E. Herrera-Viedma, F. Chiclana, and F. Herrera. A web based consensus support system for group decision making problems and incomplete preferences. Information Sciences, 180(23):4477 – 4495, 2010.

[Arr59] Kenneth J Arrow. Rational choice functions and orderings. Economica, 26(102):121–127, 1959.

[AT05] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge & Data Engineering, (6):734–749, 2005.

[Ata99] Krassimir T Atanassov. Intuitionistic fuzzy sets. In Intuitionistic fuzzy sets, pages 1–137. Springer, 1999.

[BBD+04] Antonio Bahamonde, Gustavo F Bayón, Jorge Díez, José Ramón Quevedo, Oscar Luaces, Juan José Del Coz, Jaime Alonso, and Félix Goyache. Feature subset selection for learning preferences: A case study. In Proceedings of the twenty-first international conference on Machine learning, page 7. ACM, 2004.

[BEF84] James C Bezdek, Robert Ehrlich, and William Full. Fcm: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2-3):191–203, 1984.

[BL+07] James Bennett, Stan Lanning, et al. The netflix prize. In Proceedings of KDD cup and workshop, volume 2007, page 35. New York, NY, USA., 2007.

[Bli74] Jean M Blin. Fuzzy relations in group decision theory. 1974.

[Bli76] Jean-Marie Blin. A linear assignment formulation of the multiattribute decision problem. Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle, 10(V2):21–32, 1976.

[BM03] Gustavo EAPA Batista and Maria Carolina Monard. An analysis of four missing data treatment methods for supervised learning. Applied artificial intelligence, 17(5-6):519–533, 2003.

[BM05] Jean-Pierre Brans and Bertrand Mareschal. Promethee methods. In Multiple criteria decision analysis: state of the art surveys, pages 163–186. Springer, 2005.

[Bog73] Kenneth P Bogart. Preference structures i: Distances between transitive preference relations. Journal of Mathematical Sociology, 3(1):49–67, 1973.

[Bog75] Kenneth P Bogart. Preference structures. ii: distances between asymmetric relations. SIAM Journal on Applied Mathematics, 29(2):254–262, 1975.

[Bre96] Leo Breiman. Bagging predictors. Machine learning, 24(2):123–140, 1996.

[BS97] Marko Balabanović and Yoav Shoham. Fab: content-based, collaborative recommendation. Communications of the ACM, 40(3):66–72, 1997.

[BSS78] James C Bezdek, Bonnie Spillman, and Richard Spillman. A fuzzy relation space for group decision theory. Fuzzy Sets and systems, 1(4):255–268, 1978.

[BV85] Jean-Pierre Brans and Ph Vincke. Note—a preference ranking organisation method: (the promethee method for multiple criteria decision-making). Management science, 31(6):647–656, 1985.

[CAHV09] Francisco Javier Cabrerizo, Sergio Alonso, and Enrique Herrera-Viedma. A consensus model for group decision making problems with unbalanced fuzzy linguistic information. International Journal of Information Technology & Decision Making, 8(01):109–131, 2009.

[CCK$^+$05] Michael Chau, Reynold Cheng, Ben Kao, et al. Uncertain data mining: A new research direction. In Proceedings of the Workshop on the Sciences of the Artificial, Hualien, Taiwan, pages 199–204, 2005.

[CCKN06] Michael Chau, Reynold Cheng, Ben Kao, and Jackey Ng. Uncertain data mining: An example in clustering location data. In Pacific-Asia conference on knowledge discovery and data mining, pages 199–204. Springer, 2006.

[CHVAH08] Francisco Chiclana, Enrique Herrera-Viedma, Sergio Alonso, and Francisco Herrera. Cardinal consistency of reciprocal preference relations: a characterization of multiplicative transitivity. IEEE transactions on fuzzy systems, 17(1):14–23, 2008.

[CK91] Wade D. Cook and Moshe Kress. A multiple criteria decision model with ordinal preference data. European Journal of Operational Research, 54(2):191 – 198, 1991.

[CKS86] Wade D. Cook, Moshe Kress, and Lawrence M. Seiford. Information and preference in partial orders: A bimatrix representation. Psychometrika, 51(2):197–207, Jun 1986.

[CL02] Marie Chavent and Yves Lechevallier. Dynamical clustering of interval data: Optimization of an adequacy criterion based on hausdorff distance. In Classification, clustering, and data analysis, pages 53–60. Springer, 2002.

[CLCH13] C. Chen, C. S. Lin, F. S. Chen, and W. Hung. The criteria for evaluating the effectiveness of preference aggregation methods. In 2013 International Conference on Fuzzy Theory and Its Applications (iFUZZY), pages 167–170, Dec 2013.

[CMF08] Laurent Candillier, Frank Meyer, and Françoise Fessant. Designing specific weighted similarity measures to improve collaborative filtering systems. In Industrial Conference on Data Mining, pages 242–255. Springer, 2008.

[Col13] Josep M Colomer. Ramon llull: from 'ars electionis' to social choice theory. Social Choice and Welfare, 40(2):317–328, 2013.

[CSS98] William W Cohen, Robert E Schapire, and Yoram Singer. Learning to order things. In Advances in Neural Information Processing Systems, pages 451–457, 1998.

[DB13] Sébastien Destercke and Thomas Burger. Toward an axiomatic definition of conflict between belief functions. IEEE transactions on cybernetics, 43(2):585–596, 2013.

[dCmdC85] Jean-Antoine-Nicolas de Caritat marquis de Condorcet. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. De l'Imprimerie royale, 1785.

[DDCLB08] Jorge Díez, Juan José Del Coz, Oscar Luaces, and Antonio Bahamonde. Clustering people according to their preference criteria. Expert Systems with Applications, 34(2):1274–1284, 2008.

[Dem67] Authur P. Dempster. Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Statist., 38(2):325–339, 04 1967.

[Den95] Thierry Denœux. A k-nearest neighbor classification rule based on dempster-shafer theory. IEEE transactions on systems, man, and cybernetics, 25(5):804–813, 1995.

[Den06] Thierry Denœux. The cautious rule of combination for belief functions and some extensions. In 2006 9th International Conference on Information Fusion, pages 1–8. IEEE, 2006.

[DG77] Persi Diaconis and Ronald L Graham. Spearman's footrule as a measure of disarray. Journal of the Royal Statistical Society: Series B (Methodological), 39(2):262–268, 1977.

[DJBW03] Susan Dumais, Thorsten Joachims, Krishna Bharat, and Andreas Weigend. Sigir 2003 workshop report: implicit measures of user interests and preferences. In SIGIR Forum, volume 37, pages 50–54, 2003.

[DKS15a] Thierry Denœux, Orakanya Kanjanatarakul, and Songsak Sriboonchitta. Ek-nnclus: a clustering procedure based on the evidential k-nearest neighbor rule. Knowledge-Based Systems, 88:57–69, 2015.

[DKS15b] Thierry Denœux, Orakanya Kanjanatarakul, and Songsak Sriboonchitta. Ek-nnclus: a clustering procedure based on the evidential k-nearest neighbor rule. Know.-Based Syst., 88(C):57–69, November 2015.

[DM04] Thierry Denœux and M-H Masson. Evclus: evidential clustering of proximity data. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 34(1):95–109, 2004.

[DMÖ+12] Stéphane Deparis, Vincent Mousseau, Meltem Öztürk, Christophe Pallier, and Caroline Huron. When conflict induces the expression of incomplete preferences. European Journal of Operational Research, 221(3):593–602, 2012.

[Doy04] Jon Doyle. Prospects for preferences. Computational Intelligence, 20(2):111–136, 2004.

[DP88] Didler Dubois and Henri Prade. Representation and combination of uncertainty with belief functions and possibility measures. Computational intelligence, 4(3):244–264, 1988.

[EBMBY15] Faten Elarbi, Tassadit Bouadi, Arnaud Martin, and Boutheina Ben Yaghlane. Preference fusion for community detection in social networks. In 24ème Conférence sur la Logique Floue et ses Applications, 24ème Conférence sur la Logique Floue et ses Applications, Poitiers, France, November 2015.

[Eme13] Peter Emerson. The original borda count and partial voting. Social Choice and Welfare, 40(2):353–358, 2013.

[EMS04] Zied Elouedi, Khaled Mellouli, and Philippe Smets. Assessing sensor reliability for multisensor data fusion within the transferable belief model. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 34(1):782–787, 2004.

[EMSBY14] Amira Essaid, Arnaud Martin, Grégory Smits, and Boutheina Ben Yaghlane. A Distance-Based Decision in the Credal Level. In International Conference on Artificial Intelligence and Symbolic Computation (AISC 2014), pages 147 – 156, Sevilla, Spain, December 2014.

[EMSY14] Amira Essaid, Arnaud Martin, Grégory Smits, and Boutheina Ben Yaghlane. A distance-based decision in the credal level. In International Conference on Artificial Intelligence and Symbolic Computation, pages 147–156. Springer, 2014.

[FH03] Johannes Fürnkranz and Eyke Hüllermeier. Pairwise preference learning and ranking. In European conference on machine learning, pages 145–156. Springer, 2003.

[FH10] Johannes Fürnkranz and Eyke Hüllermeier. Preference Learning, pages 789–795. Springer US, Boston, MA, 2010.

[FH11] Johannes Fürnkranz and Eyke Hüllermeier. Preference Learning: An Introduction, pages 1–17. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[FHR+14] Johannes Fürnkranz, Eyke Hüllermeier, Cynthia Rudin, Roman Slowinski, and Scott Sanner. Preference Learning (Dagstuhl Seminar 14101). Dagstuhl Reports, 4(3):1–27, 2014.

[FKM+04] Ronald Fagin, Ravi Kumar, Mohammad Mahdian, D Sivakumar, and Erik Vee. Comparing and aggregating rankings with ties. In Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pages 47–58. ACM, 2004.

[FKS03a] Ronald Fagin, Ravi Kumar, and Dakshinamurthi Sivakumar. Comparing top k lists. SIAM Journal on discrete mathematics, 17(1):134–160, 2003.

[FKS03b] Ronald Fagin, Ravi Kumar, and Dandapani Sivakumar. Efficient similarity search and classification via rank aggregation. In Proceedings of the 2003 ACM SIGMOD international conference on Management of data, pages 301–312. ACM, 2003.

[FMR05] José Figueira, Vincent Mousseau, and Bernard Roy. Electre methods. In Multiple criteria decision analysis: State of the art surveys, pages 133–153. Springer, 2005.

[FS97] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1):119–139, 1997.

[FSW+17] Daniel Faber, Jennie Stephens, Victor Wallis, Roger Gottlieb, Charles Levenstein, Patrick CoatarPeter, and Boston Editorial Group of CNS.

Trump's electoral triumph: Class, race, gender, and the hegemony of the polluter-industrial complex, 2017.

[FVJ09] P Thomas Fletcher, Suresh Venkatasubramanian, and Sarang Joshi. The geometric median on riemannian manifolds with application to robust atlas estimation. NeuroImage, 45(1):S143–S152, 2009.

[GHLR01] Cyril Goutte, Lars Kai Hansen, Matthew G Liptrot, and Egill Rostrup. Feature-space clustering for fmri meta-analysis. Human brain mapping, 13(3):165–183, 2001.

[GJ94] Zoubin Ghahramani and Michael I Jordan. Supervised learning from incomplete data via an em approach. In Advances in neural information processing systems, pages 120–127, 1994.

[GMS99a] Salvatore Greco, Benedetto Matarazzo, and Roman Slowinski. Rough approximation of a preference relation by dominance relations. European Journal of operational research, 117(1):63–83, 1999.

[GMS99b] Salvatore Greco, Benedetto Matarazzo, and Roman Slowinski. The use of rough sets and fuzzy sets in mcdm. In Multicriteria decision making, pages 397–455. Springer, 1999.

[GSB10] Shengbo Guo, Scott Sanner, and Edwin V Bonilla. Gaussian process preference elicitation. In Advances in neural information processing systems, pages 262–270, 2010.

[HA85] Lawrence Hubert and Phipps Arabie. Comparing partitions. Journal of Classification, 2(1):193–218, Dec 1985.

[HAMA16] Mohammad Haghighat, Mohamed Abdel-Mottaleb, and Wadee Alhalabi. Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition. IEEE Transactions on Information Forensics and Security, 11(9):1984–1996, 2016.

[HB01] Richard J Hathaway and James C Bezdek. Fuzzy c-means clustering of incomplete data. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 31(5):735–744, 2001.

[HG05] Hani Hamdan and Gérard Govaert. Mixture model clustering of uncertain data. In The 14th IEEE International Conference on Fuzzy Systems, 2005. FUZZ'05., pages 879–884. IEEE, 2005.

[HKBR99] Jonathan L Herlocker, Joseph A Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1999, pages 230–237. Association for Computing Machinery, Inc, 1999.

[HP08] Günter Hägele and Friedrich Pukelsheim. The electoral systems of Nicolas of Cusa in the Catholic Concordance and beyond, pages 229–249. 01 2008.

[HVCHA07] Enrique Herrera-Viedma, Francisco Chiclana, Francisco Herrera, and Sergio Alonso. Group decision-making model with incomplete fuzzy preference relations based on additive consistency. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 37(1):176–189, 2007.

[IJW+19] Eugene Ie, Vihan Jain, Jing Wang, Sanmit Navrekar, Ritesh Agarwal, Rui Wu, Heng-Tze Cheng, Morgane Lustman, Vince Gatto, Paul Covington, et al. Reinforcement learning for slate-based recommender systems: A tractable decomposition and practical methodology. arXiv preprint arXiv:1905.12767, 2019.

[JD88] Anil K Jain and Richard C Dubes. Algorithms for clustering data. Englewood Cliffs: Prentice Hall, 1988, 1988.

[JE09] Mohsen Jamali and Martin Ester. Trustwalker: a random walk model for combining trust-based and item-based recommendation. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 397–406. ACM, 2009.

[JGEB01] Anne-Laure Jousselme, Dominic Grenier, and Éloi Bossé. A new distance between two bodies of evidence. Information Fusion, 2(2):91 – 101, 2001.

[JM12] Anne-Laure Jousselme and Patrick Maupin. Distances in evidence theory: Comprehensive survey and generalizations. International Journal of Approximate Reasoning, 53(2):118–145, 2012.

[Joa02] Thorsten Joachims. Optimizing search engines using clickthrough data. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 133–142. ACM, 2002.

[JPW07] R Kang-Xing Jin, David C Parkes, and Patrick J Wolfe. Analysis of bidding networks in ebay: aggregate preference identification through community detection. 2007.

[JS18] Radim Jiroušek and Prakash P Shenoy. A new definition of entropy of belief functions in the dempster–shafer theory. International Journal of Approximate Reasoning, 92:49–65, 2018.

[Kam03a] Toshihiro Kamishima. Nantonac collaborative filtering: recommendation based on order responses. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 583–588. ACM, 2003.

[Kam03b] Toshihiro Kamishima. Nantonac collaborative filtering: Recommendation based on order responses. In Proceedings of the Ninth ACM SIGKDD

International Conference on Knowledge Discovery and Data Mining, KDD '03, pages 583–588, New York, NY, USA, 2003. ACM.

[KBGM09] Georgia Koutrika, Benjamin Bercovitz, and Hector Garcia-Molina. Flexrecs: expressing and combining flexible recommendations. In Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, pages 745–758. ACM, 2009.

[KDC18] John Klein, Sebastien Destercke, and Olivier Colot. Idempotent conjunctive and disjunctive combination of belief functions by distance minimization. International Journal of Approximate Reasoning, 92:32–48, 2018.

[Ken48] Maurice George Kendall. Rank correlation methods. 1948.

[KF87] George J. Klir and Tina A. Folger. Fuzzy Sets, Uncertainty, and Information. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1987.

[KJS96] David J Ketchen Jr and Christopher L Shook. The application of cluster analysis in strategic management research: an analysis and critique. Strategic management journal, pages 441–458, 1996.

[KKA10] Toshihiro Kamishima, Hideto Kazawa, and Shotaro Akaho. A survey and empirical comparison of object ranking methods. In Preference learning, pages 181–201. Springer, 2010.

[KKKR13] Bahador Khaleghi, Alaa Khamis, Fakhreddine O Karray, and Saiedeh N Razavi. Multisensor data fusion: A review of the state-of-the-art. Information fusion, 14(1):28–44, 2013.

[KL51] Solomon Kullback and Richard A Leibler. On information and sufficiency. The annals of mathematical statistics, 22(1):79–86, 1951.

[KM01] Slim Ben Khelifa and Jean-Marc Martel. A distance-based collective weak ordering. Group Decision and Negotiation, 10(4):317–329, 2001.

[KP13] Noam Koenigstein and Ulrich Paquet. Xbox movies recommendations: variational bayes matrix factorization with embedded feature selection. In Proceedings of the 7th ACM Conference on Recommender Systems, pages 129–136. ACM, 2013.

[KS63] John G. Kemeny and James Laurie Snell. Preference ranking: An axiomatic approach. In Mathematical models in the social sciences, chapter 2. Blaisdell, New York, 1963.

[LCLM16] Z. Li, R. Chen, L. Liu, and G. Min. Dynamic resource discovery based on preference and movement pattern similarity for large-scale social internet of things. IEEE Internet of Things Journal, 3(4):581–589, Aug 2016.

[LDSS04] Dan Li, Jitender Deogun, William Spaulding, and Bill Shuart. Towards missing data imputation: a study of fuzzy k-means clustering method. In International Conference on Rough Sets and Current Trends in Computing, pages 573–579. Springer, 2004.

[LDSS05] Dan Li, Jitender Deogun, William Spaulding, and Bill Shuart. Dealing with missing data: Algorithms based on fuzzy set and rough set theories. In James F. Peters and Andrzej Skowron, editors, Transactions on Rough Sets IV, pages 37–57, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[Lip13] David Lippman. Voting theory. In Math in society. David Lippman, 2013.

[LR19] Roderick JA Little and Donald B Rubin. Statistical analysis with missing data, volume 793. John Wiley & Sons, 2019.

[LS02] Risto Lahdelma and Pekka Salminen. Pseudo-criteria versus linear utility function in stochastic multi-criteria acceptability analysis. European Journal of Operational Research, 141(2):454–469, 2002.

[Luc59] R Duncan Luce. Individual choice behavior. 1959.

[Luc12] R Duncan Luce. Individual choice behavior: A theoretical analysis. Courier Corporation, 2012.

[M+67] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[Mar12] Arnaud Martin. About conflict in the theory of belief functions. In Belief Functions: Theory and Applications, pages 161–168. Springer, 2012.

[Mar19] Arnaud Martin. Conflict management in information fusion with belief functions. In Information Quality in Information Fusion and Decision Making, pages 79–97. Springer, 2019.

[MC11] José M. Merigó and Montserrat Casanovas. Decision-making with distance measures and induced aggregation operators. Computers & Industrial Engineering, 60(1):66–76, 2011.

[MCG+99] Tim Miranda, Mark Claypool, Anuja Gokhale, Tim Mir, Pavel Murnikov, Dmitry Netes, and Matthew Sartin. Combining content-based and collaborative filters in an online newspaper. In In Proceedings of ACM SIGIR Workshop on Recommender Systems. Citeseer, 1999.

[MD08] Marie-Hélène Masson and Thierry Denœux. Ecm: An evidential version of the fuzzy c-means algorithm. Pattern Recognition, 41(4):1384–1397, 2008.

[MGL13] José M Merigó and Anna M Gil-Lafuente. Induced 2-tuple linguistic generalized aggregation operators and their application in decision-making. Information Sciences, 236:1–16, 2013.

[MJO08] Arnaud Martin, Anne-Laure Jousselme, and Christophe Osswald. Conflict measure for the discounting operation on belief functions. In 2008 11th International Conference on Information Fusion, pages 1–8, June 2008.

[MO07] Arnaud Martin and Christophe Osswald. Toward a combination rule to deal with partial conflict and specificity in belief functions theory. In 2007 10th International Conference on Information Fusion, pages 1–8. IEEE, 2007.

[NKC+06] Wang Kay Ngai, Ben Kao, Chun Kit Chui, Reynold Cheng, Michael Chau, and Kevin Y Yip. Efficient clustering of uncertain data. In Sixth International Conference on Data Mining (ICDM'06), pages 436–445. IEEE, 2006.

[ÖTV05] Meltem Öztürke, Alexis Tsoukiàs, and Philippe Vincke. Preference Modelling, pages 27–59. Springer New York, New York, NY, 2005.

[Paz99] Michael J Pazzani. A framework for collaborative, content-based and demographic filtering. Artificial intelligence review, 13(5-6):393–408, 1999.

[PC09] Seung-Taek Park and Wei Chu. Pairwise preference regression for cold-start recommendation. In Proceedings of the third ACM conference on Recommender systems, pages 21–28. ACM, 2009.

[pCqX19] Hui ping Chen and Gui qiong Xu. Group decision making with incomplete intuitionistic fuzzy preference relations based on additive consistency. Computers & Industrial Engineering, 135:560 – 567, 2019.

[PKCK12] Deuk Hee Park, Hyea Kyeong Kim, Il Young Choi, and Jae Kyeong Kim. A literature review and classification of recommender systems research. Expert Systems with Applications, 39(11):10059 – 10072, 2012.

[Pla75] Robin L Plackett. The analysis of permutations. Journal of the Royal Statistical Society: Series C (Applied Statistics), 24(2):193–202, 1975.

[PLP11] Yu Peng, Qinghua Luo, and Xiyuan Peng. Uck-means: A customized k-means for clustering uncertain measurement data. In 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), volume 2, pages 1196–1200. IEEE, 2011.

[RD76] Douglas W Rae and Hans Daudt. The ostrogorski paradox: a peculiarity of compound majority decision. European Journal of Political Research, 4(4):391–398, 1976.

[Rou87] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20:53 – 65, 1987.

[Roy68] Bernard Roy. Classement et choix en présence de points de vue multiples. Revue française d'informatique et de recherche opérationnelle, 2(8):57–75, 1968.

[RS93] Bernard Roy and R Slowinski. Criterion of distance between technical programming and socio-economic priority. RAIRO-Operations Research, 27(1):45–60, 1993.

[Saa02] TL Saaty. Decision making with the analytic hierarchy process. Scientia Iranica, 9(3):215–229, 2002.

[Sha76] Glenn Shafer. A Mathematical Theory of Evidence. Princeton University Press, Princeton, 1976.

[SM95] Upendra Shardanand and Pattie Maes. Social information filtering: algorithms for automating "word of mouth". In Chi, volume 95, pages 210–217. Citeseer, 1995.

[Sme90] Philippe Smets. The combination of evidence in the transferable belief model. IEEE Trans. Pattern Anal. Mach. Intell., 12(5):447–458, May 1990.

[Sme93] Philippe Smets. Belief functions: the disjunctive rule of combination and the generalized bayesian theorem. International Journal of approximate reasoning, 9(1):1–35, 1993.

[Sme00] Philippe Smets. Data fusion in the transferable belief model. In Proceedings of the third international conference on information fusion, volume 1, pages PS21–PS33. IEEE, 2000.

[Sme07] Philippe Smets. Analyzing the combination of conflicting belief functions. Information fusion, 8(4):387–412, 2007.

[Ste05] Theodor J. Stewart. Dealing with Uncertainties in MCDA, pages 445–466. Springer New York, New York, NY, 2005.

[Sug09] Jeffrey Sugerman. Using the disc® model to improve communication effectiveness. Industrial and Commercial Training, 41(3):151–154, 2009.

[Tan84] Tetsuzo Tanino. Fuzzy preference orderings in group decision making. Fuzzy Sets and Systems, 12(2):117 – 131, 1984.

[Tar72] Robert Tarjan. Depth first search and linear graph algorithms. SIAM JOURNAL ON COMPUTING, 1(2), 1972.

[THJA04] Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. Support vector machine learning for interdependent and structured output spaces. In Proceedings of the twenty-first international conference on Machine learning, page 104. ACM, 2004.

[Tho53] Robert L Thorndike. Who belongs in the family? Psychometrika, 18(4):267–276, 1953.

[VdCL00] Rosanna Verde, Francisco de AT de Carvalho, and Yves Lechevallier. A dynamical clustering algorithm for multi-nominal data. In Data analysis, classification, and related methods, pages 387–393. Springer, 2000.

[Via15] Paolo Viappiani. Characterization of scoring rules with distances: application to the clustering of rankings. In Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.

[WD91] Michael P. Wellman and Jon Doyle. Preferential semantics for goals. In In Proceedings of the National Conference on Artificial Intelligence, pages 698–703, 1991.

[WLXC05] David Williams, Xuejun Liao, Ya Xue, and Lawrence Carin. Incomplete-data classification using logistic regression. In Proceedings of the 22nd International Conference on Machine learning, pages 972–979. ACM, 2005.

[WWY15] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. Collaborative deep learning for recommender systems. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, pages 1235–1244. ACM, 2015.

[Xu07] Zeshui Xu. Intuitionistic preference relations and their application in group decision making. Information sciences, 177(11):2363–2379, 2007.

[Xu15] Zeshui Xu. Uncertain multi-attribute decision making: Methods and applications. Springer, 2015.

[Yag87] Ronald R Yager. On the dempster-shafer framework and new combination rules. Information sciences, 41(2):93–137, 1987.

[Yag88] Ronald R Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on systems, Man, and Cybernetics, 18(1):183–190, 1988.

[Yag03] Ronald R Yager. Induced aggregation operators. Fuzzy sets and systems, 137(1):59–69, 2003.

[YDP99] C-H Yeh, H. Deng, and H. Pan. Multi-criteria analysis for dredger dispatching under uncertainty. Journal of the Operational Research Society, 50(1):35–43, 1999.

[YF98] Ronald R Yager and Dimitar Filev. Operations for granular computing: mixing words and numbers. In 1998 IEEE International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98CH36228), volume 1, pages 123–128. IEEE, 1998.

[YF99] Ronald R Yager and Dimitar P Filev. Induced ordered weighted averaging operators. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 29(2):141–150, 1999.

[YST+13] Jia-Ching Ying, Bo-Nian Shi, Vincent S Tseng, Huan-Wen Tsai, Kuang Hung Cheng, and Shun-Chieh Lin. Preference-aware community detection for item recommendation. In 2013 Conference on Technologies and Applications of Artificial Intelligence, pages 49–54. IEEE, 2013.

[Zad65] Lotfi A Zadeh. Fuzzy sets. Information and control, 8(3):338–353, 1965.

[ZBM17] Yiru Zhang, Tassadit Bouadi, and Arnaud Martin. Preference fusion and condorcet's paradox under uncertainty. In 2017 20th International Conference on Information Fusion (Fusion), pages 1–8. IEEE, 2017.

[ZBM18a] Yiru Zhang, Tassadit Bouadi, and Arnaud Martin. A clustering model for uncertain preferences based on belief functions. In International Conference on Big Data Analytics and Knowledge Discovery, pages 111–125. Springer, 2018.

[ZBM18b] Yiru Zhang, Tassadit Bouadi, and Arnaud Martin. An empirical study to determine the optimal k in ek-nnclus method. In International Conference on Belief Functions, pages 260–268. Springer, 2018.

[ZC03] Dao-Qiang Zhang and Song-Can Chen. Clustering incomplete data using kernel-based fuzzy c-means algorithm. Neural processing letters, 18(3):155–162, 2003.

[ZKL15] Hongyi Zhang, Irwin King, and Michael R Lyu. Incorporating implicit link preference into overlapping community detection. In Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

[ZMP17] Kuang Zhou, Arnaud Martin, and Quan Pan. Evidence combination for a large number of sources. In 2017 20th International Conference on Information Fusion (Fusion), pages 1–8. IEEE, 2017.

[ZMP18] Kuang Zhou, Arnaud Martin, and Quan Pan. A belief combination rule for a large number of sources. Journal of Advances in Information Fusion, 13(2), 2018.

[ZWFM06] Sheng Zhang, Weihong Wang, James Ford, and Fillia Makedon. Learning from incomplete ratings using non-negative matrix factorization. In Proceedings of the 2006 SIAM international conference on data mining, pages 549–553. SIAM, 2006.

[ZYST19] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. Deep learning based recommender system: A survey and new perspectives. ACM Computing Surveys (CSUR), 52(1):5, 2019.

List of Figures

UNIVERSITÉ DE RENNES 1

Titre : Modélisation et gestion des préférences imparfaites avec la théorie des fonctions de croyance

Mot clés : préférences imparfaites ; agrégation de préférences ; apprentissage des préférences ; théorie des fonctions de croyance ;

Résumé : La modélisation et gestion de préférences ouvrent de nouveaux défis, surtout avec l'émerge de service dans le monde numérique. Ces travaux se concentrent sur les imperfections dans l'information des préférences, telles que l'incertitude, l'imprécision et l'incomplétude. Dans cette thèse, nous passons en revue les méthodes existantes sur l'agrégation et l'apprentissage des préférences. Fondé sur la théorie des fonctions de croyance, nous proposons une modèle permettant à raisonner les préférences au niveau du coupe à partir d'un degré de croyance.
Ce modèle est capable de représenter l'incertitude, l'imprécision ainsi que l'incomplétude par l'ignorance totale dans le cadre des fonctions de croyance. Nous proposons ensuite des stratégies pertinentes pour fusionner de multiple préférences crédibilistes. De plus, une distance sur les préférences imparfaites,nommée Weighted Singleton Distance (WSD), est introduite afin de tenir compte différemment des quatre types de relations de préférence.
La classification non-supervisée sur les préférences crédibilistes est aussi étudiée en distinguant les préférences complètes et incomplètes, avec une théorème d'impossibilité sur classification des objects crédibilists proposée et prouvée.

Title: Modeling and management of imperfect preferences with the theory of belief functions

Keywords: imperfect preferences, preference aggregation, preference learning, theory of belief functions

Abstract: With the emergence service in the digital world, modeling and managing preferences bring new challenges. This work focuses on imperfections in preference information, such as uncertainty, imprecision and incompleteness. In this thesis, we review state-of-the-art methods on preference aggregation and preference learning. Based on the theory of belief functions, we propose a model of preference information on the pairs of alternatives (or objects) being compared.
This model is capable of expressing uncertainty, imprecision and as incompleteness through total ignorance in the framework of the theory of belief functions. We then propose relevant strategies to fuse multiple belief preferences. In addition, a novel distance, named Weighted Singleton Distance (WSD), on imperfect preferences is introduced to take into account the four types of preference relationships differently.
The unsupervised classification on imperfect preferences with BFpref model is also studied by distinguishing complete and incomplete preferences, with an impossibility theorem proposed and proved.